

GI-19745307 264 ----- TWW KERKITYFKLYRQLGEKEVAVDDAELK ----- QINS
 ORF84 WO 2006/078318 TWW KERKITYFKLYRQLGEKEVAVDDAELK ----- QINSEGOPCT/US2005/027239
 GI-28810263 264 ----- TWDAPKERKITYFKLYRQLGEKEVAVDDAELK ----- QINSEGQ
 GI-21909640 150 ----- TWDAPKERKITYFKLYRQLGEKEVAVDDAELK ----- QINSEGQ
 GI-19224141 601 PHSSVKVEANKEVTIMNHNETLTFSGKNIWENDREDORPAKIQVQLQNGOKMPNQIQEV

GI-19745307 305 ----- QEISVTVTNQLVTDEKGMRITYISVKEVDK ----- NGELLEPKD ----- YIKKED
 ORF84 305 ----- QEISVTVTNQLVTDEKGMRITYISVKEVDK ----- NGELLEPKD ----- YIKKED
 GI-28810263 305 ----- QEISVTVTNQLVTDEKGMRITYISVKEVDK ----- NGELLEPKD ----- YIKKED
 GI-21909640 191 ----- QEISVTVTNQLVTDEKGMRITYISVKEVDK ----- NGELLEPKD ----- YIKKED
 GI-19224141 661 TKDNEWSYHEKDLPKYDANNQEYKYSVEVNVPDGYKVSYLGNEFTNTRETEFVEHQNNF

GI-19745307 349 GLTVTNTYVKPTSG HYDIEVTFC ----- NGHIDITEDTPDIVSGENQMK
 ORF84 349 GLTVTNTYVKPTSG HYDIEVTFC ----- NGHIDITEDTPDIVSGENQMK
 GI-28810263 349 GLTVTNTYVKPTSG HYDIEVTFC ----- NGHIDITEDTPDIVSGENQMK
 GI-21909640 235 GLTVTNTYVKPTSG HYDIEVTFC ----- NGHIDITEDTPDIVSGENQMK
 GI-19224141 721 NLEFGNAEINGQSGSKIIDEDTLTSFKGKKIWKNDAENRPQAIVQOLYADGVAVEGOTK

GI-19745307 394 QIEGEDS ----- KPIDEVT ----- ENNLIEFGKNTMPGEE
 ORF84 394 QIEGEDS ----- KPIDEVT ----- ENNLIEFGKNTMPGEE
 GI-28810263 394 QIEGEDS ----- KPIDEVT ----- ENNLIEFGKNTMPGEE
 GI-21909640 280 QIEGEDS ----- KPIDEVT ----- ENNLIEFGKNTMPGEE
 GI-19224141 781 EISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPGYDVTTYSANDITNTKREVITQQG

GI-19745307 424 ----- DGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATIHKFSKRD
 ORF84 424 ----- DGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATIHKFSKRD
 GI-28810263 424 ----- DGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATIHKFSKRD
 GI-21909640 310 ----- DGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATIHKFSKRD
 GI-19224141 841 PKLEIEETLPLESGASGGTTIEDSRPVDTLSGLSSEQGQSGDMTIEEDSATIHKFSKRD

GI-19745307 473 IDGKELAGATMELRDSGGNTISTWISDGQVDFYLMPGKYTFVETAAPDGYEIATAITFT
 ORF84 473 IDGKELAGATMELRDSGGNTISTWISDGQVDFYLMPGKYTFVETAAPDGYEIATAITFT
 GI-28810263 473 IDGKELAGATMELRDSGGNTISTWISDGQVDFYLMPGKYTFVETAAPDGYEIATAITFT
 GI-21909640 359 IDGKELAGATMELRDSGGNTISTWISDGQVDFYLMPGKYTFVETAAPDGYEIATAITFT
 GI-19224141 901 IDGKELAGATMELRDSGGNTISTWISDGQVDFYLMPGKYTFVETAAPDGYEIATAITFT

GI-19745307 533 VNEQGQVTVNGKATKGDAHIVIVDAYKPTKGSGQVIDIEEKLPEOGHSGSTTEIEDSKS
 ORF84 533 VNEQGQVTVNGKATKGDAHIVIVDAYKPTKGSGQVIDIEEKLPEOGHSGSTTEIEDSKS
 GI-28810263 533 VNEQGQVTVNGKATKGDAHIVIVDAYKPTKGSGQVIDIEEKLPEOGHSGSTTEIEDSKS
 GI-21909640 419 VNEQGQVTVNGKATKGDAHIVIVDAYKPTKGSGQVIDIEEKLPEOGHSGSTTEIEDSKS
 GI-19224141 961 VNEQGQVTVNGKATKGDAHIVIVDAYKPTKGSGQVIDIEEKLPEOGHSGSTTEIEDSKS

GI-19745307 593 SDVIIGGQG -----
 ORF84 593 SDVIIGGQG -----
 GI-28810263 593 SDVIIGGQGEVVDTTEDTOSGMTGHS -----
 GI-21909640 479 SDVIIGGQGEVVDTTEDTOSGMTGHSSTTEIEDSKSSDVIIIGQGEVVDTTEDTOSGMT
 GI-19224141 1021 SDLIIGGQGEVVDTTEDTOSGMTGHS -----

GI-19745307 602 ----- DIVETTEDTGTGMHDSGCITEVEDTKLVQSFHF DNK
 ORF84 602 ----- DIVETTEDTGTGMHDSGCITEVEDTKLVQSFHF DNK
 GI-28810263 619 ----- GSTTKEEDSKSSDVIVCGQGQIVETTEDTGTGMHDSGCITEVEDTKLVQSFHF DNK
 GI-21909640 539 GHSGSTTKEEDSKSSDVIVCGQGQIVETTEDTGTGMHDSGCITEVEDTKLVQSFHF DNK
 GI-19224141 1047 ----- GSTTEIEDSKSSDVIIIGQGQIVETTEDTGTGMHDSGCITEVEDTKLVQSFHF DNK

GI-19745307 639 ESESNSEIPKKDKPKSNTSLPATGEKQHNMFIVMTCSLTSSVFVISLTKRLSSC
 ORF84 639 ESESNSEIPKKDKPKSNTSLPATGEKQHNMFIVMTCSLTSSVFVISLTKRLSSC
 GI-28810263 676 EPESENSEIPKKDKPKSNTSLPATGEKQHNMFIVMTCSLTSSVFVISLTKRLSSC
 GI-21909640 599 EPESENSEIPKKDKPKSNTSLPATGEKQHNMFIVMTCSLTSSVFVISLKSRLSSC
 GI-19224141 1104 EPESENSEIPKKDKPKSNTSLPATGEKQHNMFIVMTCSLTSSVFVISLKSRLSSC

FIGURE 54A

GI-19224WO 2006/078318-MKRNKLLATAILATALGTASLNQNVKAETAGVUSSGOLTIKNSITPCT/US2005/027239

ORF80 PCT¹ 1 MERERMKRNKLLATAILATALGTASLNQNVKAETAGVUSSGOLTIKNSITPCT/US2005/027239
 GI-21909636 1 MERERMKRNKLLATAILATALGTASLNQNVKAETAGVUSSGOLTIKNSITPCT/US2005/027239
 GI-28810259 1 MERERMKRNKLLATAILATALGTASLNQNVKAETAGVUSSGOLTIKNSITPCT/US2005/027239
 GI-19745303 1 MERERMKRNKLLATAILATALGTASLNQNVKAETAGVUSSGOLTIKNSITPCT/US2005/027239
 GI-13621428 1 MERERMKRNKLLATAILATALGTASLNQNVKAETAGVUSSGOLTIKNSITPCT/US2005/027239

GI-19224137 56 PNTDYTFSVNPDSAPTCESN-LETRPGTAVN-NODIKVSYNTDKTSQNEKQVVADFMK
 ORF80 60 PTAETFTIIPDMTASGREGS-LEIINNGIVEGLDKQVIVKYKNTDKTSQNTKIAQDFDFSK
 GI-21909636 56 PRADYTFKVEADSTASGKTKDGLEIKPGIVNGLT-EQIISYTNTDNDPSIVKSTEFDPSK
 GI-28810259 61 PRADYTFKVEADSTASGKTKDGLEIKPGIVNGLT-EQIISYTNTDNDPSIVKSTEFDPSK
 GI-19745303 56 PRADYTFKVEADDNAKGKTKDGLEIKPGIVDGLENKTIIHYCNSDKTTANEKSVNFDFAN
 GI-13621428 48 PNTDFTFNIIEPDITVNEDGNKFKG-----VALNTPMTKVITYTNSDNGGSNTKTAEFDSE

GI-19224137 114 VTFPSVGTIYRYVVTENKGTAE-GVTYDDTKILVDVYVGN--NEKGCGLEPKYIVSNKEDSA
 ORF80 119 VKFFEAIGVYRYMVKNDKD-GLTYCDKKITVDVYVGNKANNEBGFELVYIVSKECTSS
 GI-21909636 115 VVFPGIGVYRYTVSEKQGDVE-GITYDTKKIITVDVYVGN--KEGGGFEPKIEIVSKEDQTD
 GI-28810259 120 VVFPGIGVYRYTVSEKQGDVE-GITYDTKKWTVDVYVGN--KEGGGFEPKIEIVSKEDQTD
 GI-19745303 116 VKFPQGVYVRYTVSEVNGNNA-GIAYDQQQNTVDVYVGN--REDGGFEAKYIVSTECCQS
 GI-13621428 103 VTEEKPGVYVYKVTEEKIDKVPGVSYDTTSVTVQHILWLN-EFQQKPVATIYIVGYKEGS

GI-19224137 171 TMEPIOFNNSFETTSLSKIEEVGTGNTGDKHAFTFTLTLQNEYEAASSVVKIEENCO--
 ORF80 178 TRKPKIPTFNSLKKTSLKIEKQITGNAQDRKASFNFTLTLQSEYNTGSVVKIEQDGS--
 GI-21909636 172 VNKPKVNFNNSFATTSLKVKKNVSGNTGELQKEFDFTLTLQNESTNEKNDQIVSLOKGN--
 GI-28810259 177 VNKPKVNFNNSFATTSLKVKKNVSGNTGELQKEFDFTLTLQNESTNEKNDQIVSLOKGN--
 GI-19745303 173 DKKPKVFKNEFDTTSLSKVKRVTGNTGEHORSPSFTLIIPNECHENGQVNILOGGB--
 GI-13621428 161 KVPIOKNSLSDSTTIVKPKVSGTGCDRSKDFNFCITLKANQYXRASEKWMIENTTKGG

GI-19224137 229 ----TKDVKIGEAVKETLNDOSVILSKLPFGINYKVBEAEANOGGYTTATLNDG--EK
 ORF80 236 ----KKDWIGTPIKFTLGKGSVMSKLPIGINYVMSSEDEANKDGYTTATLNEOGKEK
 GI-21909636 230 ----KEVKIGTPYKEELKNGESTGDLKLPVGINYVNEEANNDGYTTASLNEG-DGO
 GI-28810259 235 ----KEVKIGTPYKEELKNGESTGDLKLPVGINYVNEEANNDGYTTASLNEG-DGO
 GI-19745303 231 ----TAKVIGEPPFTLKDQESVILSKLPVGIBYRVTEEDVHNDGVTSATLNDG-E-V
 GI-13621428 220 QAPVQTBASTDOEVHFTLKDGESIKVTLNLPGVDTWVTEDDYKSEKTTINVEVSPQDGAV

GI-19224137 283 LSTVNLG-OEHKTDKTADEIVVTNNRDTQVPTGVVGT LAPFAVL SIVAIGGVIVITKRN
 ORF80 292 SSELFLSTONOKTDESAD EIVVTNKRD TOVPTGVVGT LAPFAVL SIVAIGGVIVITKRN
 GI-21909636 285 SKMYOOLD-MEOKTDESAD EIVVTNKRD TOVPTGVVGT LAPFAVL SIVAIGGVIVITKRN
 GI-28810259 290 SKMYOOLD-MEOKTDESAD EIVVTNKRD TOVPTGVVGT LAPFAVL SIVAIGGVIVITKRN
 GI-19745303 285 TDCYMLG-DSKQITDKSDEIVVTNKRD TOVPTGVVGT LAPFAVL SIVAIGGVIVITKRN
 GI-13621428 280 KNTAGNSTEQETSTDKDMTIPETNKKDIEVPTGVAMTVAPYALGIVAVGCALYEVNNKHN

GI-19224137 342 A
 ORF80 352 A
 GI-21909636 344 A
 GI-28810259 349 A
 GI-19745303 344 A
 GI-13621428 340 A

173/487

PCT/US2005/027239

GI-2190WO 2006/078318

GI-28810261	1	MLFSVVMILTMALAFNQTVLAKDSTV
GI-19224139	1	MLFSVVMILTMALAFNQTVLAKDSTV
ORF82	1	LLFQRVKIFLLTIVLSLSVLFKNNERRRLRKYWKMLFSVVMILTMALAFNQTVLAKDSTV
GI-19745305	1	MRKYWKMLFSVVMILTMALAFNQTVLAKDSTV

GI-21909638	20	QTSISVENVLERAGDSTPFSTIALESIDAMKTIEEITIAGSGKASFSPLTFTTVGQTYRV
GI-28810261	26	QTSISVENVLERAGDSTPFSTIALESIDAMKTIEEITIAGSGKASFSPLTFTTVGQTYRV
GI-19224139	26	QTSISVENVLERAGDSTPFSTIALESIDAMKTIEEITIAGSGKASFSPLTFTTVGQTYRV
ORF82	61	QTSISVENVLERAGDSTPFSTIALESIDAMKTIDEITIAGSGKASFSPLTFTTVGQTYRV
GI-19745305	32	QTSISVENVLERAGDSTPFSTIALESIDAMKTIDEITIAGSGKASFSPLTFTTVGQTYRV

GI-21909638	80	YQKPSQNKDQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRNLVKPIP
GI-28810261	86	YQKPSQNKDQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRNLVKPIP
GI-19224139	86	YQKPSQNKDQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRNLVKPIP
ORF82	121	YQKPSQNKDQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRNLVKPIP
GI-19745305	92	YQKPSQNKDQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRNLVKPIP

GI-21909638	140	PRQPNIPKTPLEPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-28810261	146	PRQPNIPKTPLEPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-19224139	146	PRQPNIPKTPLEPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
ORF82	181	PRQPNIPKTPLEPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-19745305	152	PRQPDIPKTPLEPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

FIGURE 56

PCT/US2005/027239 174/487

!!SEQUENCE_LIST 1.0

(Peptide) FASTA of: gi-50913505.pep from: 1 to: 1036 September 15, 2004 18:46

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences

z-scores computed from opt scores

z-score	obs (=)	exp (*)
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	5	5:====*
48	10	5:====*=====
50	12	4:====*=====
52	4	4:====*
54	6	3:==*====
56	4	3:==*==
58	5	2:==*==
60	3	2:==*==
62	0	1:*
64	1	1:*
66	1	1:*
68	0	1:*
70	1	1:*
72	0	0:
74	0	0:
76	1	0:=
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	1	0:=
98	0	0:

FIGURE 57

PCT/US2005/027239
 100 0 0:
 102 0 0:
 104 0 0:
 106 1 0:=
 108 0 0:
 110 0 0:
 112 0 0:
 114 0 0:
 116 0 0:
 118 0 0:
 >120 1 0:=

Joining threshold: 39, opt. threshold: 27, opt. width: 16, reg.-scaled

The best scores are:

init1 initn opt z-sc E(54)..

/home/morama/gas/pili/align/gi-50913505.pep		Begin: 1 End: 1036	
! gi 50913505 ref YP_059477.1 Collag...	6697	6697 3452.1	9.6e-189
/home/morama/gas/pili/align/gi-19224141.pep		Begin: 48 End: 144	
! gi 19224141 gb AAL86412.1 AF447492...	63	100 159 105.9	0.023
/home/morama/gas/pili/align/gi-21909640.pep		Begin: 147 End: 449	
! gi 21909640 ref NP_663908.1 protei...	35	35 136 96.3	0.08
/home/morama/gas/pili/align/gi-13621428.pep		Begin: 57 End: 318	
! gi 13621428 gb AAK33238.1 hypothet...	33	33 91 75.6	1.1
/home/morama/gas/pili/align/gi-50913506.pep		Begin: 33 End: 428	
! gi 50913506 ref YP_059478.1 Fimbri...	70	149 86 71.3	1.9
/home/morama/gas/pili/align/gi-13621432.pep		Begin: 14 End: 56	
! gi 13621432 gb AAK33241.1 conserve...	40	65 78 68.0	2.9
/home/morama/gas/pili/align/gi-19745301.pep		Begin: 241 End: 466	
! gi 19745301 ref NP_606437.1 putati...	52	52 73 64.8	4.3
/home/morama/gas/pili/align/gas15.pep		Begin: 492 End: 739	
! GAS15 GAS15	43	68 69 61.4	6.6
/home/morama/gas/pili/align/gi-21909636.pep		Begin: 176 End: 298	
! gi 21909636 ref NP_663904.1 conser...	31	31 62 60.8	7.1
/home/morama/gas/pili/align/gi-28810259.pep		Begin: 181 End: 303	
! gi 28810259 gb BAC63197.1 hypoth...	31	31 62 60.7	7.2
/home/morama/gas/pili/align/gi-19224139.pep		Begin: 90 End: 143	
! gi 19224139 gb AAL86410.1 AF447492...	43	43 54 58.9	8.9
/home/morama/gas/pili/align/gi-19745305.pep		Begin: 96 End: 149	
! gi 19745305 ref NP_606441.1 hypoth...	43	43 54 58.8	9
/home/morama/gas/pili/align/orf82.pep		Begin: 125 End: 178	
! TRANSLATE of: orf82.seq check: 4296...	43	43 54 58.2	9.6
/home/morama/gas/pili/align/gi-21909638.pep		Begin: 84 End: 137	
! gi 21909638 ref NP_663906.1 hypoth...	43	43 52 58.0	9.9
\End of List			

gi-50913505.pep

/home/morama/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

SCORES Init1: 6697 Initn: 6697 Opt: 6697 z-score: 3452.1 E(): 9.6e-189
 >>/home/morama/gas/pili/align/gi-50913505.pep (1036 aa)
 initn: 6697 init1: 6697 opt: 6697 Z-score: 3452.1 expect(): 9.6e-189
 Smith-Waterman score: 6697; 100.0% identity in 1036 aa overlap
 (1-1036:1-1036)

10 20 30 40 50 60
 gi-50913505. MYSRLKRELVIVINRKKYKLIRLMVTVGQLIFSQLVLPPIRRLGLQMISTQTKVIPQEIVT

FIGURE 57A

PC T11H99E51P7P3P

gi-50913505. MYSRLKRELVIVINRKKYKLIRLMVTVGLIFSQLVLPIRRGLQMI
10 20 30 40 50 60

gi-50913505. QTETQGTQVVATKQKLESENSSLKVALKRESGFEHNATIDASLDTE
70 80 90 100 110 120

gi-50913505. QTETQGTQVVATKQKLESENSSLKVALKRESGFEHNATIDASLDTE
70 80 90 100 110 120

gi-50913505. VTMALLELRKQGLSIVDTKIVRIQSSTNQRNDITTLTFKNGLSLEG
130 140 150 160 170 180

gi-50913505. VTMALLELRKQGLSIVDTKIVRIQSSTNQRNDITTLTFKNGLSLEG
130 140 150 160 170 180

gi-50913505. VNPNDTVQTITPTIKQDADGKVNLVFTGRLGKQVIIIVSTTRLKEE
190 200 210 220 230 240

gi-50913505. VNPNDTVQTITPTIKQDADGKVNLVFTGRLGKQVIIIVSTTRLKEE
190 200 210 220 230 240

gi-50913505. GAVGLSQDKDRPPYSKPITVNILKPKLSSIESLDSKDFEIVKTID
250 260 270 280 290 300

gi-50913505. GAVGLSQDKDRPPYSKPITVNILKPKLSSIESLDSKDFEIVKTID
250 260 270 280 290 300

gi-50913505. SKQYEVLKTDYQSAKDSTPQTRDILFGEYTVPLVMNKGHNN
310 320 330 340 350 360

gi-50913505. SKQYEVLKTDYQSAKDSTPQTRDILFGEYTVPLVMNKGHNN
310 320 330 340 350 360

gi-50913505. AAPALIQPRSFRSLTPRSTRMKRSAPVEKFEGELEHHKRIDL
370 380 390 400 410 420

gi-50913505. AAPALIQPRSFRSLTPRSTRMKRSAPVEKFEGELEHHKRIDL
370 380 390 400 410 420

gi-50913505. HDTSDLYRLYLDMTGKKNPLDILVVVDKSGSMQE
430 440 450 460 470 480

gi-50913505. HDTSDLYRLYLDMTGKKNPLDILVVVDKSGSMQE
430 440 450 460 470 480

gi-50913505. TFDYSSYQGESFNRCQIHYRYRGIVSVSDGIRRDAVKNSLLGV
490 500 510 520 530 540

gi-50913505. TFDYSSYQGESFNRCQIHYRYRGIVSVSDGIRRDAVKNSLLGV
490 500 510 520 530 540

gi-50913505. SVIGFQGSADYHAGKWP
550 560 570 580 590 600

gi-50913505. SVIGFQGSADYHAGKWP
550 560 570 580 590 600

gi-50913505. GTNYHAALLKAKEILNEVKDDGRRKIMIF
610 620 630 640 650 660

gi-50913505. GTNYHAALLKAKEILNEVKDDGRRKIMIF
610 620 630 640 650 660

FIGURE 57B

PCT/US2005/027239

gi-50913505. QEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSEEHYYGITDAALE
 ||||| ||||| ||||| ||||| ||||| |||||
 gi-50913505. QEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSEEHYYGITDAALE
 670 680 690 700 710 720

gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
 ||||| ||||| ||||| ||||| ||||| |||||
 gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
 730 740 750 760 770 780

gi-50913505. DIIDKVVFTPCKTTSQPKGVTLTFKSDYKVDDETYTLISFNVKASDEAYEKYKDNEGRYS
 ||||| ||||| ||||| ||||| |||||
 gi-50913505. DIIDKVVFTPCKTTSQPKGVTLTFKSDYKVDDETYTLISFNVKASDEAYEKYKDNEGRYS
 790 800 810 820 830 840

gi-50913505. EMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLKYKHPVIQVKTVPIFTKVD
 ||||| ||||| ||||| ||||| |||||
 gi-50913505. EMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLKYKHPVIQVKTVPIFTKVD
 850 860 870 880 890 900

gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYLYETKAKLGYTLP
 ||||| ||||| ||||| ||||| |||||
 gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYLYETKAKLGYTLP
 910 920 930 940 950 960

gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIIVGSMTA
 ||||| ||||| ||||| |||||
 gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIIVGSMTA
 970 980 990 1000 1010 1020

gi-50913505. TVALLFYRRQHRKKQY
 ||||| |||||
 gi-50913505. TVALLFYRRQHRKKQY
 1030

gi-50913505.pep
 /home/morama/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492_9 protein F2 [Streptococcus pyogenes]

SCORES : Init1: 63 Initn: 100 Opt: 159 z-score: 105.9 E(): 0.023
 >>/home/morama/gas/pili/align/gi-19224141.pep (1161 aa)
 initn: 100 init1: 63 opt: 159 z-score: 105.9 expect(): 0.023
 Smith-Waterman score: 159; 36.7% identity in 98 aa overlap
 (895-990:48-144)

gi-50913505. SDASVNYMADGREQKLKYKHPVIQVKTVPIFTKVDADNNQKKLAGVEFELRKEDKK-IV
 :| :|:|:| | | :| | | :| |
 gi-19224141. FILGLLLFIGLGSVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKGT
 20 30 40 50 60 70

930 940 950 960 970 980

FIGURE 57C

PC-T/US005/Z7339

gi-50913505. WEKGTTGSNGQLNFYLYQKGTYYLYETKAKLGYTLPEVAVANNGDIK-VKHPIEG
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 gi-19224141. VQTFTSNDKGIVDAQNLQPG-TYTLKEETAPDGYDKTSRTWTVIVYENGYTKLVENPYNG
 80 90 100 110 120 130

990 1000 1010 1020 1030
 gi-50913505. ELKSKDGSYMIKNYKTYQLPSSGGRGSQIFIIIVGSMTATVALLFYRRQHRKKQY
 | : | | : |
 gi-19224141. EIISKAGSKDVSSLQLENPKMSVVSKEYGKTEVSSGAADFYRNHAAYFKMSFELKQKDKS
 140 150 160 170 180 190

gi-50913505.pep

/home/morrama/gas/pili/align/gi-21909640.pep

gi|21909640|ref|NP_663908.1| protein F2 like fibronectin-binding protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 35 InitN: 35 Opt: 136 z-score: 96.3 E(): 0.08
 >>/home/morrama/gas/pili/align/gi-21909640.pep (656 aa)
 initn: 35 init1: 35 opt: 136 z-score: 96.3 expect(): 0.08
 Smith-Waterman score: 148; 24.5% identity in 339 aa overlap
 (686-1005:147-449)

660 670 680 690 700 710
 gi-50913505. NVTRSQEGLKLAIDEFKARYPNLSTIYSLGVSKDINSDTASSSPVVL----KYLSGEEHYY
 : | | : | : : | : : | : | : | : | : | : | : | : | : | : | : | : |
 gi-21909640. TNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEV-
 120 130 140 150 160 170

720 730 740 750 760 770
 gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDLSQYVVDYYDKQPDVLTFRSKVNDTEILYQ
 : | : | : | : : : : | : | : | : | : | : | : | : | : | : | : | : |
 gi-21909640. -AVDDAELQ----INSEGQQEISVTWT-NQLVT--DEKGMAITYSVKEVDKNGELLEP
 180 190 200 210 220

780 790 800 810 820 830
 gi-50913505. KDQVQ-EAGKDIIDKVVFTPCKTTSQPKGVTLTFKSDYKVDEYTYTLSFNVKASDEAYE
 | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 gi-21909640. KDYIKKEDGLTVNTYV--KPTSG-HYDIEVTFGNGH-ID--TEDTTTPDIVSGEN---
 230 240 250 260 270

840 850 860 870
 gi-50913505. KYKDNEGRYSEMGSSTDYGTNQTSSGKGGLPSNSDA-SVNY-MADGR-----
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 gi-21909640. QMKQIEGEDSKPIDEVTE--NNLIEFGKNTMPGEEDGTNSNKYEVEDSRPVDTLSGLSS
 280 290 300 310 320 330

880 890 900 910 920 930
 gi-50913505. EQKLKYKHPVIQVKTPITFTKVDADNNQQKLAGVEFLRKEDKKIVWEKGTTGSNGQIN
 | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 gi-21909640. EQGQSGDMTIEEDSATHLKFSKRDTDG--KELAGATMELRDSSGKTI--STWISDGQVK
 340 350 360 370 380

940 950 960 970 980 990
 gi-50913505. FKYLQKGKTYLYETKAKLGYTLPEVAVANNGDIKVKHPIEGELKSKDGSYMIKNY
 | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 gi-21909640. DFYLMMPGK-YTFVETAAPDGY-----EVATAITFTVNEQQVTVNGKATKGDAAHIVMV
 390 400 410 420 430 440

1000 1010 1020 1030
 gi-50913505. KIYQLPSSGGRGSQIFIIIVGSMTATVALLFYRRQHRKKQY

FIGURE 57D

gi-21909640. DAYK-PTKGSGQVIDIEEKLPEQGHSGSTETEDSKSSDVIIIGGQGEVVDTTEDTQSGM
450 460 470 480 490 500

gi-50913505.pep

/home/morama/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

```
SCORES Init1: 33 Initn: 33 Opt: 91 z-score: 75.6 E(): 1.1  
>/home/morama/gas/pili/align/gi-13621428.pep (340 aa)  
initn: 33 init1: 33 opt: 91 z-score: 75.6 expect(): 1.1  
Smith-Waterman score: 95; 19.9% identity in 271 aa overlap  
(568-819:57-318)
```

540	550	560	570	580	590
gi-50913505.	NKLSVIGFQGSADYHAGKWPDQSPRGFYQPQNLNNNSRDAELLKGWSTNSLLDPNTLTAL				
			: : : : : : : : : :		
gi-13621428.	VNGAKLTVTKNLSDLVNSNALIPMTDFTFKIEPDPTVNEDGNKFKGVALNTPMTKVTYTMS				
			30 40 50 60 70 80		

	660	670	680	690	700
gi-50913505.	SNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTAS-	SSPVVLKVYLSGEEH			
	:	:	:	:	:
gi-13621428.	KPVATYIVGYKEGSVKPI-QFKN--	SLDSTTLTVKKVSGTGGDRSKDFNFGTLKANQ			
	150	160	170	180	190
					200

710	720	730	740	750		
gi-50913505.	YYGITDTAELEKTLN---	KIVEDSKLSQL---	GISDSDLSQLYVDYYDKQPDVLVT	--R		
	:: : : . : . : ; : : : . : : :					
gi-13621428.	YYKASEKVMIIEKTTKGQAPVQTEASIDQLYHFTLKDGESIKVTNLPGVGDYVVTEDDYK					
	210	220	230	240	250	260

	760	770	780	790	800	810
gi-50913505	KSKVNDETEIILYQKDQVQEAGKDEIIDKVVFPTKTTSQPKGVTLTFKSDYKVVDDEYTYTL	:	:	:	:	:
gi-13621428	SEKYTTNVEVSPQDGAVRNIAQMSSTEQETSTDKDMT	—	—	—	TTFTNKKDFEVPTGVAMTV	
	270	280	290	300	310	

820 830 840 850 860 870
gi-50913505. SFNVKASDEAYEKYKDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQ
:
gi-13621428. APYTALGIVAVGGALYFVKKKNA
320 330 340

gi-50913505.pep

/home/morama/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes MGAS10394].

SCORES Init1: 70 Initn: 149 Opt: 86 z-score: 71.3 E(): 1.9
>>/home/morama/gas/pili/align/gi-50913506.pep (556 aa)

FIGURE 57E

initn: 149 initl: 70 opt: 86 z-score: 71.3 expect(): 1.9
 Smith-Waterman score: 120; 21.5% identity in 469 aa overlap
 (503-966:33-428)

480	490	500	510	520	
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDAAVKNSLLGVNG-----L					
:: :: : :: : ::					
gi-50913506. NRRETVREKILITAKKMLACLAIALAVVGLGMTRVS-ALSKDDAQLKITNIEGGPTVTL					
10	20	30	40	50	60
530	540	550	560	570	580
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWPQSPRGFYQPMLNNSRDAELLKGWSTNS					
: : : : : : : : : : : :					
gi-50913506. YKIGEGVYNTNGDSFINFK---YAEGVSLTETGPTSQEIT-TIANGINTGKIKPFSSEN					
70	80	90	100	110	
590	600	610	620	630	640
gi-50913505. LLDPNTLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFTISDGVPFTYFGEDGYRSG					
: :: : :: : : : : : : : :					
gi-50913506. VSISNGTATYNARGASVYIALLTGAT-----DGRTYNPILLAAS----YNGEGNLVTK					
120	130	140	150	160	
650	660	670	680	690	700
gi-50913505. NGSSNDRNNVTRSQECKLAIDEFKARYPNLSTYSLGVSKDINSDTASSSPVVLKYLSGE					
: :: : :: : : : : : : :					
gi-50913506. NIDS--KSNLYLGQ--TSVA---KSSLPSITKKVTGTTIDDVNKKTTSLGSVLSYSLTFE					
170	180	190	200	210	
710	720	730	740	750	760
gi-50913505. EHYYGITDTAELEKTLNKIVEDSKLSQLGIISDLSQYVDYDKQPDVLVTRKSKVNDTE					
: :: : : : : : : : : :					
gi-50913506. LPSY-----TKEAVNKTVY-----VSDNMSEGLTF--NFNSLTVEWKGMANITE					
220	230	240	250	260	
770	780	790	800	810	820
gi-50913505. ILYQKDQVQEAGKDIIDKVVFPTKTTSQPKGKVTLTFSYKVDDEYTYSFNVKASDE					
:: : : : : : : :					
gi-50913506. -----DCSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----					
270	280	290			
830	840	850	860	870	880
gi-50913505. AYEKYKDNEGRYSEMGSDDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVI					
: : :: : : : : : : : :					
gi-50913506. SYKAVVNNKAIVGEECNPNKAFFYSNNPTKGNTYDNLDDKKPDK-GNGITSKEDSK					
300	310	320	330	340	350
890	900	910	920	930	940
gi-50913505. QVKTVPLFTKVDADNNQKKLAGVEFELRKEDKKIVWEKGTTGNGQLNFKYLOKGKTYY					
: : : : : : : : : :					
gi-50913506. IVYTYQIAFRKVDS-VSKTPPLIGAIFGVYDTSNKL-LDIVTINKNGYAISTQVSSGK-YK					
360	370	380	390	400	
950	960	970	980	990	1000
gi-50913505. LYETKAKLGYTLPENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIIQLPSSGR					
: : : : :					
gi-50913506. IKELKAPKGYSLNTETYEITANWVTATVKTSAKSSTTYTSDKNKATDNSEQVGWLKNGI					
410	420	430	440	450	460

gi-50913505.pep

/home/morama/gas/pili/align/gi-13621432.pep

gi|13621432|gb|AAK33241.1| conserved hypothetical protein [Streptococcus pyogene]

FIGURE 57F

PCT/US2005/027239 181/487

s]

SCORES Init1: 40 Initn: 65 Opt: 78 z-score: 68.0 E(): 2.9
 >>/home/morama/gas/pili/align/gi-13621432.pep (450 aa)
 initn: 65 init1: 40 opt: 78 z-score: 68.0 expect(): 2.9
 Smith-Waterman score: 78; 37.0% identity in 46 aa overlap
 (368-411:14-56)

	340	350	360	370	380	390
gi-50913505.	KGHNNTINIIYIRSTRPLGLKPIGAAPALIQPRSFRSLTPRSTRMK--RSAPVEKFEGELE					
gi-13621432.	MTRTNYQKKRMTCPVETEDITYRRKKIKGRRQAILAQFEPELV					
	10	20	30	40		

	400	410	420	430	440	450
gi-50913505.	HHKRIDYLGDNQNPNPTTIDDKDEHDTSDLYRLYLDLMTGKKNPLDIEVVVDKSGSMQEG					
gi-13621432.	HHELI---GDSCTCPDCHGTLTEIGSVVQRQELVFIAPAQLKRINHVQHAYKCQTCSDNSL					
	50	60	70	80	90	100

gi-50913505.pep

/home/morama/gas/pili/align/gi-19745301.pep

gi|19745301|ref|NP_606437.1| putative collagen binding protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 52 Initn: 52 Opt: 73 z-score: 64.8 E(): 4.3
 >>/home/morama/gas/pili/align/gi-19745301.pep (524 aa)
 initn: 52 init1: 52 opt: 73 z-score: 64.8 expect(): 4.3
 Smith-Waterman score: 95; 23.7% identity in 245 aa overlap
 (759-989:241-466)

	730	740	750	760	770	780
gi-50913505.	DSKLSQLGISDSLSQLYDVKQPDVLVTRKSKVNDETEILYQKDQVQEAG-KDIIDKVV					
gi-19745301.	ETIDPDFNEGKEIKYTHILGADLFYSANNPRASTNDE--LLSQVKVLEKGYRD--DSTT					
	220	230	240	250	260	

	790	800	810	820	830	840
gi-50913505.	FTPKTTSQPKGVTLT--FKSDYKVDD-EYTYTLSFNVKASDEAYEKYKDNEGRYSEM					
gi-19745301.	YANLTSVEFRAATOLAIYYFTDSVLDNLADYHGFALTTEALNATKETVAYAEDRANLP					
	270	280	290	300	310	320

	850	860	870	880	890	
gi-50913505.	GDSDTDY--GTNOTSS-GKGGLPSNSDAVNMYADGREQKLKYKHPVIQVKTVPTIFT					
gi-19745301.	NISNLDFYVPNSNKYQSLIGTOYHP-ESLVDLIERMEDKQAPIIPTITHKLTIISKTVTGTI-					
	330	340	350	360	370	380

	900	910	920	930	940	950
gi-50913505.	KVDADNNQQKKLAGVFELRKEDKKIVWEKGTTGSN-GQLNFKYLQKGK-TYYLYETKARL					
gi-19745301.	---AD--KKKEFNFEIHLKSSDGQAI--SGTYPTNSGELT---VTDGKATFTLKDGESLI					
	390	400	410	420	430	

	960	970	980	990	1000	1010
gi-50913505.	GYTLPEN-PWEVAVANNGDIKVHKPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFII					
gi-19745301.	VEGLPSGYSYEITETGASDYEV--VNGK-NAPDGKATKASVKEDETITFENRKDLVPPT					

FIGURE 57G

PCT/US05/27239

440 450 460 470 480 490

1020 1030

gi-50913505. VGSMTATVALLFYRRQHRKKQY

gi-19745301. GLTTDGAIYLWLLLIVLLGLWWLIGRKGLKND

500 510 520

gi-50913505.pep

/home/morama/gas/pili/align/gas15.pep

GAS15 GAS15

SCORES Init1: 43 Initn: 68 Opt: 69 z-score: 61.4 E(): 6.6
 >>/home/morama/gas/pili/align/gas15.pep (762 aa)
 initn: 68 init1: 43 opt: 69 z-score: 61.4 expect(): 6.6
 Smith-Waterman score: 100; 21.4% identity in 252 aa overlap
 (641-873:492-739)

620 630 640 650 660
 gi-50913505. AKELILNEVKDDGRRKIMIFISDGVPFTFYFGEDGYRSGNGSSNDRNNVTRSQ--EGSKLAI
 |||| :|:: : ::::|::| ::::||
 gas15.pep HTAGRDLFKYTVKPRDTDPDTFLKHKKVIEKGYRE-KGQALEYESGLTETQLRAATQLAI
 470 480 490 500 510 520

670 680 690 700 710 720
 gi-50913505. DEF--KARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSEEHYYGITDTAELEKTLNKI
 ||| :|: : :: :| :| :| :| :| :| :| :| :| :| :| :|
 gas15.pep YYFTDSAELDKDLKDYHGFGDMNDSTLAVAKILVEY-AQDSNPQQLTDLDFFIPNNNNKY
 530 540 550 560 570

730 740 750 760 770
 gi-50913505. VEDSKLSQLGISDSLSQLYVDDYDKQPDVLVT----RKSKVN---DETETLYQKDQVQEAE
 :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
 gas15.pep --QSLIGTQWHPEDLVDIIRMEDKKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNN
 580 590 600 610 620 630

780 790 800 810 820 830
 gi-50913505. GKDIIIDKVVFTPPTISQPK-GKVTLTFKSDYKVDDE-YTYTLSFNVKASDEAYEKYKDNE
 :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
 gas15.pep KQELLSOTVKTDKTNLEFKDGKATINLKHEGESLTQQLPEGYSYLVKETDSEGYKVVKVNS
 640 650 660 670 680 690

840 850 860 870 880 890
 gi-50913505. GRYSEMGSDDTDYGTNQT---SSGKGGLPSNSDASVN-YMADGREQKLPHYKHPVIQVKT
 :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
 gas15.pep QEVANATVSKTGITSDETAFENNKEPVVPTGVDQKINGYLALIVIAGISLGIWGIHTIR
 700 710 720 730 740 750

900 910 920 930 940 950
 gi-50913505. VPITFTKVDADNNQKIKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYLYET
 gas15.pep IRKHD
 760

gi-50913505.pep

/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.8 E(): 7.1
 >>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)
 initn: 31 init1: 31 opt: 62 z-score: 60.8 expect(): 7.1
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap
 (181-305:176-298)

	160	170	180	190	200	210
gi-50913505.	DITTTLTFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGR					
	150	160	170	180	190	
gi-21909636.	VDVYVGNKEGGGFEPKFIIVSKEQGTDVKKPVNFMNSFATTSLVKKNVSGN-----TGE					
	220	230	240	250	260	
gi-50913505.	LGKQ----VIIVSTTRLKEEQTISLD SYGELVIDGAVGLSQKDRPPYSKPITVNILKPKL					
	200	210	220	230	240	250
gi-21909636.	LQKEFDFTTLNESTNFKKDQIVSLQK-GNEKFEVKIGTPYFKFLKGESIQLDKLPVGI					
	270	280	290	300	310	320
gi-50913505.	SSIESSLDSKDFEIVKTIDNLYTWDDQ--FYLLDFISKQYEVLKTDYQSAKDSTPQTRDI					
	260	270	280	290	300	310
gi-21909636.	TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESADEIVVTNKRDTQVPTGV					
	330	340	350	360	370	380
gi-50913505.	LFGEYTVEPLVMNKGHNNNTINIIYIRSTRPLGLKPIGAAPALIQPRSFRSLTPRSTRMKRS					
gi-21909636.	VGTLAPFAVL SIVAIGGVIYITKRKKA					
	320	330	340			

gi-50913505.pep

/home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.7 E(): 7.2
 >>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)
 initn: 31 init1: 31 opt: 62 z-score: 60.7 expect(): 7.2
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap
 (181-305:181-303)

	160	170	180	190	200	210
gi-50913505.	DITTTLTFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGR					
	160	170	180	190	200	
gi-28810259.	VDVYVGNKEGGGFEPKFIIVSKEQGTDVKKPVNFMNSFATTSLVKKNVSGN-----TGE					
	220	230	240	250	260	
gi-50913505.	LGKQ----VIIVSTTRLKEEQTISLD SYGELVIDGAVGLSQKDRPPYSKPITVNILKPKL					
	210	220	230	240	250	260
gi-28810259.	LQKEFDFTTLNESTNFKKDQIVSLQK-GNEKFEVKIGTPYFKFLKGESIQLDKLPVGI					
	270	280	290	300	310	320
gi-50913505.	SSIESSLDSKDFEIVKTIDNLYTWDDQ--FYLLDFISKQYEVLKTDYQSAKDSTPQTRDI					
	270	280	290	300	310	320
gi-28810259.	TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESADEIVVTNKRDTQVPTGV					

FIGURE 57I

PCT/US05/237239

330 340 350 360 370 380

gi-50913505. LFGEYTVEPLVMNKGNNTINIIYIRSTRPLGLKPIGAAPALIQPRSFRSLTPRSTRMKRS

gi-28810259. VGTLAPFAVLSIVAIIGGVIYITKRKKA

330 340

gi-50913505.pep

/home/morama/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.9 E(): 8.9
 >>/home/morama/gas/pili/align/gi-19224139.pep (189 aa)

initn: 43 init1: 43 opt: 54 z-score: 58.9 expect(): 8.9

Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
 (742-796:90-143)

720 730 740 750 760

gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDLSQYVVDYYDKQP--DVLVTRKSKVNDETEIL

|| || || : || |

gi-19224139. ITIAGSGKASFSPLTFTTVGQTYRKYQKPSQNKDYQADTTVFDVLVYV--TYDEDGTL

60 70 80 90 100 110

770 780 790 800 810 820

gi-50913505. YQKDQVQEAGKDIIDKVVFPTPKTTSQPKGVTLTFKSDYKVDDETYTLFSNVKASDEAY

| : || : : || | : |

gi-19224139. VAKVISRAGDEEKSAITFKPKRLVKPIPQPNIPKTPPLPLAGEVKSLGILSIVLLGL

120 130 140 150 160 170

gi-50913505.pep

/home/morama/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP_606441.1| hypothetical protein [Streptococcus pyogenes MGAS82 32]

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.8 E(): 9
 >>/home/morama/gas/pili/align/gi-19745305.pep (195 aa)

initn: 43 init1: 43 opt: 54 z-score: 58.8 expect(): 9

Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
 (742-796:96-149)

720 730 740 750 760

gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDLSQYVVDYYDKQP--DVLVTRKSKVNDETEIL

|| || || : || |

gi-19745305. ITIAGSGKASFSPLTFTTVGQTYRKYQKPSQNKDYQADTTVFDVLVYV--TYDEDGTL

70 80 90 100 110 120

770 780 790 800 810 820

gi-50913505. YQKDQVQEAGKDIIDKVVFPTPKTTSQPKGVTLTFKSDYKVDDETYTLFSNVKASDEAY

| : || : : || | : |

gi-19745305. VAKVISRAGDEEKSAITFKPKRLVKPIPQPDIPKTPPLPLAGEVKSLGILSIVLLGL

130 140 150 160 170 180

gi-50913505.pep

/home/morama/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672
 generated symbols 1 to: 224.

PCT/US05/237239
GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.2 E(): 9.6
>>/home/morama/gas/pili/align/orf82.pep (224 aa)
initn: 43 init1: 43 opt: 54 Z-score: 58.2 expect(): 9.6
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
(742-796:125-178)

	720	730	740	750	760	
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVVDYYDKQP--	DVLVTRKSKVNDTEIL				
orf82.pep	ITIAGSGKASFPLTFTTVGQTYTYRKYQKPSQNQDYQADTTVFDVLVYV--	TYDEDGTL				
	100	110	120	130	140	
					150	
	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFTPCKTSQPKGVTLTFKSDYKVDDETYTLFSFNVKASDEAY					
orf82.pep	VAKVISRRAGDEEKSATTFKPKRLVVKPIPPRQPNIPKTPPLPLAGEVKSLLGILSIVLLGL					
	160	170	180	190	200	210

gi-50913505.pep

/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP_663906.1| hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 43 Initn: 43 Opt: 52 z-score: 58.0 E(): 9.9
>>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)
initn: 43 init1: 43 opt: 52 Z-score: 58.0 expect(): 9.9
Smith-Waterman score: 52; 31.6% identity in 57 aa overlap
(742-796:84-137)

	720	730	740	750	760	
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVVDYYDKQP--	DVLVTRKSKVNDTEIL				
gi-21909638.	ITIAGSGKASFPLTFTTVGQTYTYRKYQKPSQNQDYQADTTVFDVLVYV--	TYDEDGTL				
	60	70	80	90	100	110
	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFTPCKTSQPKGVTLTFKSDYKVDDETYTLFSFNVKASDEAY					
gi-21909638.	VAKVISRRAGDEEKSATTFKPKRLVVKPIPPRQPNIPKTPPLPLAGEVKSLLGILSIVLLGL					
	120	130	140	150	160	170

! Distributed over 1 thread.

! Start time: Wed Sep 15 18:46:18 2004

! Completion time: Wed Sep 15 18:46:21 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:02.2

! Total CPU time: 0:00:02.3

! Output File: gi-50913505.fasta

PCT/US05/27239
SEQUENCE LIST 1.0

(Peptide) FASTA of: gi-50913506.pep from: 1 to: 556 September 15, 2004 18:45

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes MGAS10394]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:Each histogram symbol represents 1 search set sequences
z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	1	5:= *
48	8	5:====*==
50	24	5:====*=====
52	2	4:== *
54	3	3:==*
56	1	3:= *
58	2	2:=*
60	3	2:=*=
62	0	2:= *
64	2	1:=*=
66	2	1:=*=
68	1	1:=*
70	2	1:=*=
72	0	0:
74	3	0:====
76	1	0:=
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:

FIGURE 58

100	0	0:
102	0	0:
104	0	0:
106	0	0:
108	0	0:
110	0	0:
112	0	0:
114	0	0:
116	0	0:
118	0	0:
>120	1	0:=

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are:

init1 initn opt z-sc E(55) ..

/home/morrama/gas/pili/align/gi-50913506.pep	Begin: 1	End: 556	
! gi 50913506 ref YP_059478.1 Fimbri...	3454	3454	3454 1016.2 4.7e-53
/home/morrama/gas/pili/align/orf84.pep	Begin: 316	End: 567	
! TRANSLATE of: orf84.seq check: 7868...	57	83	135 75.1 1.2
/home/morrama/gas/pili/align/gi-19745307.pep	Begin: 316	End: 567	
! gi 19745307 ref NP_606443.1 protei...	57	83	135 75.1 1.2
/home/morrama/gas/pili/align/gi-21909640.pep	Begin: 202	End: 524	
! gi 21909640 ref NP_663908.1 protei...	56	81	134 75.0 1.2
/home/morrama/gas/pili/align/gi-28810263.pep	Begin: 316	End: 638	
! gi 28810263 dbj BAC63201.1 protein...	56	82	134 74.7 1.3
/home/morrama/gas/pili/align/orf80.pep	Begin: 49	End: 352	
! TRANSLATE of: orf80.seq check: 9824...	45	69	113 70.8 2.1
/home/morrama/gas/pili/align/gi-19224137.pep	Begin: 25	End: 342	
! gi 19224137 gb AAL86408.1 AF447492...	45	69	109 69.8 2.4
/home/morrama/gas/pili/align/gi-19224141.pep	Begin: 277	End: 645	
! gi 19224141 gb AAL86412.1 AF447492...	73	73	118 68.9 2.7
/home/morrama/gas/pili/align/gi-21909636.pep	Begin: 44	End: 344	
! gi 21909636 ref NP_663904.1 conser...	45	98	96 66.1 3.8
/home/morrama/gas/pili/align/gi-28810259.pep	Begin: 49	End: 349	
! gi 28810259 dbj BAC63197.1 hypothe...	45	98	96 66.0 3.8
/home/morrama/gas/pili/align/gas15.pep	Begin: 222	End: 470	
! GAS15..GAS15	42	68	96 63.8 5
/home/morrama/gas/pili/align/gi-13621428.pep	Begin: 17	End: 340	
! gi 13621428 gb AAK33238.1 hypothet...	41	41	87 63.6 5.2
/home/morrama/gas/pili/align/gi-19224135.pep	Begin: 193	End: 462	
! gi 19224135 gb AAL86406.1 AF447492...	41	41	86 61.0 7
/home/morrama/gas/pili/align/gi-50913505.pep	Begin: 503	End: 966	
! gi 50913505 ref YP_059477.1 Collag...	70	149	86 60.1 7.8
/home/morrama/gas/pili/align/gi-13621430.pep	Begin: 60	End: 143	
! gi 13621430 gb AAK33240.1 hypothet...	43	67	67 59.2 8.7
/home/morrama/gas/pili/align/gi-19745303.pep	Begin: 44	End: 344	
! gi 19745303 ref NP_606439.1 hypothet...	51	106	69 58.4 9.5

\End of List

gi-50913506.pep

/home/morrama/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes MGAS10394]

SCORES Init1: 3454 Initn: 3454 Opt: 3454 z-score: 1016.2 E(): 4.7e-53
 >>/home/morrama/gas/pili/align/gi-50913506.pep (556 aa)
 initn: 3454 init1: 3454 opt: 3454 Z-score: 1016.2 expect(): 4.7e-53
 Smith-Waterman score: 3454; 100.0% identity in 556 aa overlap

FIGURE 58A

PCT/US05/27239
(1-556:1-556)

	10	20	30	40	50	60
gi-50913506.	MTNRRET	VREKILITAKK	IMLACLA	LAVVGLGMTR	VSALS	KDDTAQLKITNIEGGPTVT
gi-50913506.	MTNRRET	VREKILITAKK	IMLACLA	LAVVGLGMTR	VSALS	KDDTAQLKITNIEGGPTVT
	10	20	30	40	50	60
	70	80	90	100	110	120
gi-50913506.	LYKIGEGVYNTNGDS	FINFKYAEGVSL	TETGPTSQEIT	TIANGINTGKIKPF	STENVSIS	
gi-50913506.	LYKIGEGVYNTNGDS	FINFKYAEGVSL	TETGPTSQEIT	TIANGINTGKIKPF	STENVSIS	
	70	80	90	100	110	120
	130	140	150	160	170	180
gi-50913506.	NGTATYNARGASVYI	ALLTGATDGR	TYNPILLAASYN	GEGLVTKNIDS	SKSNLYGQTSV	
gi-50913506.	NGTATYNARGASVYI	ALLTGATDGR	TYNPILLAASYN	GEGLVTKNIDS	SKSNLYGQTSV	
	130	140	150	160	170	180
	190	200	210	220	230	240
gi-50913506.	AKSSLP	SITKKVTGTIDDV	NKTTSLGSVL	SYSLT	TFELPSYTKEAVNKTVY	VSDNMSEGL
gi-50913506.	AKSSLP	SITKKVTGTIDDV	NKTTSLGSVL	SYSLT	TFELPSYTKEAVNKTVY	VSDNMSEGL
	190	200	210	220	230	240
	250	260	270	280	290	300
gi-50913506.	TFNFNSL	TVEWKGMAN	ITEDG	SVMVENT	KIGIAKEVN	NGFNL
gi-50913506.	TFNFNSL	TVEWKGMAN	ITEDG	SVMVENT	KIGIAKEVN	NGFNL
	250	260	270	280	290	300
	310	320	330	340	350	360
gi-50913506.	AVVNNKAIV	GEEGNPNKAEFFY	SNNPTKG	NTYDNLD	KKPDKGNGITS	SKEDSKIVYTYQIA
gi-50913506.	AVVNNKAIV	GEEGNPNKAEFFY	SNNPTKG	NTYDNLD	KKPDKGNGITS	SKEDSKIVYTYQIA
	310	320	330	340	350	360
	370	380	390	400	410	420
gi-50913506.	FRKVDSVSK	TPLIGAIFGVYDT	SNKLIDIV	FTNKNGYAI	STQVSSGKY	KIKELKAPKGYS
gi-50913506.	FRKVDSVSK	TPLIGAIFGVYDT	SNKLIDIV	FTNKNGYAI	STQVSSGKY	KIKELKAPKGYS
	370	380	390	400	410	420
	430	440	450	460	470	480
gi-50913506.	LNTETYEITANWV	TATVKTSANSK	TTTYS	SDKNKATD	NEQVGWL	KNGTFY
gi-50913506.	LNTETYEITANWV	TATVKTSANSK	TTTYS	SDKNKATD	NEQVGWL	KNGTFY
	430	440	450	460	470	480
	490	500	510	520	530	540
gi-50913506.	DVKEAYTESTKAL	TDTFSKSNEGSGT	VILLE	DIPNTKL	GELPSTG	SIGTYLFKAIGSA
gi-50913506.	DVKEAYTESTKAL	TDTFSKSNEGSGT	VILLE	DIPNTKL	GELPSTG	SIGTYLFKAIGSA
	490	500	510	520	530	540
	550					
gi-50913506.	AMIGAIGIYIV	KRRKA				
gi-50913506.	AMIGAIGIYIV	KRRKA				
	550					

FIGURE 58B

189/487

PCT/US2005/027239
gi-50913506.pep

/home/morama/gas/pili/align/orf84.pep

TRANSLATE of: orf84.seq check: 7868 from: 1 to: 2088
generated symbols 1 to: 696.

GETSEQ from morama, September 13, 2004 17:07.

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2
 >>/home/morama/gas/pili/align/orf84.pep (696 aa)
 initn: 83 init1: 57 opt: 135 z-score: 75.1 expect(): 1.2
 Smith-Waterman score: 146; 24.4% identity in 262 aa overlap
 (232-462:316-567)

gi-50913506. 210 220 230 240 250
 KTTSLGSVLSYSLTFELPSYTKEAVNKTVVYSDNMSEGLTFNFNSLTVEWKGMAN---

orf84.pep EKEAVVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAVIYSVKEVDKNGELLEPKDY
 290 300 310 320 330 340

gi-50913506. 260 270 280 290 300 310
 ITEDGSVMVENTKIGIAKEVNNGFNLSFIY---DSLESISPNIKYAVVNNKAIVGEE
 orf84.pep IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNGHIDITEDTPDI-VSGENQMKQIEGED
 350 360 370 380 390

gi-50913506. 320 330 340 350
 GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNGITTSKEDSKTVYTYQ
 orf84.pep SKPIDEVTEENLIEFGKNTMPGEEDGTSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIE
 400 410 420 430 440 450

gi-50913506. 360 370 380 390 400 410
 -----IAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAIISTQVSSGKYKIKE
 orf84.pep EDSATHIKFSKRD-IDGKELAGATMELRDSGGKTIS--TWISDGQVKDFYLMPGKYTFVE
 460 470 480 490 500 510

gi-50913506. 420 430 440 450 460 470
 LKAPKGYSLNTE-TYEITAN-WVTATVKTTSANSKSTTYTSDKNKATDNSEQVGWLKNIF
 orf84.pep TAAPDGYEIATAITFTVNEQQVTVNGKATKGDAHTVMV-DAYKPTKGSGQVEDIEEKLP
 520 530 540 550 560 570

gi-50913506. 480 490 500 510 520 530
 YSIDSRPTGNDVKEAYIESTKALTDGTTESKSNEGSGTVILLETDIPNTKLGELPSTGSTG
 orf84.pep DEQGHSGSTTEIEDSKSSDVLIIGGQGQIVETTEDQTGMHGDSGCKTEVEDTKLVQSFHF
 580 590 600 610 620 630

gi-50913506.pep

/home/morama/gas/pili/align/gi-19745307.pep

gi|19745307|ref|NP_606443.1| protein F2-like protein [Streptococcus pyogenes MGA S8232]

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2
 >>/home/morama/gas/pili/align/gi-19745307.pep (696 aa)
 initn: 83 init1: 57 opt: 135 z-score: 75.1 expect(): 1.2
 Smith-Waterman score: 146; 24.4% identity in 262 aa overlap
 (232-462:316-567)

	210	220	230	240	250	
gi-50913506.	KTTSLGSVLSYSLTFELPSYTKEAVNKTVVSDNMSEGLT	FNFNSLTVWKGKMAN				
gi-19745307.	EKEVAVDDAELKQINSEGQQEI	SVTWTNQLVTDE	--KG	MAYIYSVKEVDKNGELLEPKDY		
	290	300	310	320	330	340
	260	270	280	290	300	310
gi-50913506.	ITEDGSVMVENTKIGIAKEVNNGFNLSFIY		--DSLESISPNTSYKAVVNNKAIVGEE			
gi-19745307.	: :: : : : :: :	: : : :: : :				
	350	360	370	380	390	
	320	330	340	350		
gi-50913506.	GNP-----NKAEFFYSNNP		--TKGNNTYDNL-DKKP-DKGNGITSKEDSKIVYTYQ			
gi-19745307.	: :: : : : : : : : :					
	400	410	420	430	440	450
	360	370	380	390	400	410
gi-50913506.	-----IAFRKVDSVKSTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE					
gi-19745307.	:: : : : : : : : :					
	460	470	480	490	500	510
	420	430	440	450	460	470
gi-50913506.	LKAPKGYSLNTE-TYEITAN-WVTATVKTTSANSKTTYTSDKNKATDNSEQVGWLKNGIF					
gi-19745307.	: : : : : : : : : : : : :					
	520	530	540	550	560	570
	480	490	500	510	520	530
gi-50913506.	YSIDSRPTGNDVKEAYTESTKALTDGTTFSKSNEGSGT	VVLLETDIPNTKLGELPSTGSIG				
gi-19745307.	DEQGHSGSTTEIEDSKSSDVII	JGGQQQIVETTEDTQTCMHDGSGCKTEVEDT	KLVQSFHF			
	580	590	600	610	620	630

gi-50913506.pep
/home/moraima/gas/pili/align/gi-21909640.pep

gi|21909640|ref|NP_663908.1| protein F2 like fibronectin-binding protein [Streptococcus pyogenes MGAS315].

SCORES: Init1: 56 Initn: 81 Opt: 134 z-score: 75.0 E(): 1.2
->/home/morama/gas/pili/align/gi-21909640.pep (656 aa)
initn: 81 init1: 56 opt: 134 z-score: 75.0 expect(): 1.2
Smith-Waterman score: 156; 23.9% identity in 347 aa overlap
(232-547:202-524)

210	220	230	240	250	
gi-50913506.	KTTSLGSVLSYSLTFELPSYTKEAVNKTVYVSDNMSEGLTFNFNSLTVEWKGMAN				
	: : : : :				
gi-21909640.	EKEVAVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAITYSVIKEVDKNGELLEPKDY				
	180 190 200 210 220				
260	270	280	290	300	310
gi-50913506.	ITEDGSVMVENTKIGIAKEVNNGFNLNSFIY-----DSLESISPNIKYKAVVNNKAIVGEE				
	: : : : : : : : : : :				
gi-21909640.	IKKEDGLTVNTYV--KPTSGHYDIEVTFGNGHIDITEDTPDI-VSGENQMKQIEGED				
	230 240 250 260 270 280				

FIGURE 58D

PCT/US05/27239

gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKDP-DKGNGITSKEDSKIVYTYQ
 ::| | || ::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |:
 gi-21909640. SKPIDEVTEENNLLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSEQQQSGDMTIE
 290 300 310 320 330 340

gi-50913506. 360 370 380 390 400 410
 -----IAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
 ||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |:
 gi-21909640. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWLSDGQVKDFYLMPGKYTFVE
 350 360 370 380 390 400

gi-50913506. 420 430 440 450 460 470
 LKAPKGYSLNTE-TYEITAN-WVTATVKTANSKSTTYTSKDNKATDNSEQVGWLKNGIF
 ||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |:
 gi-21909640. TAAPDGYESATAITFTVNEQQQVTVNNGATKGDAHIVMV-DAYKPTKGSGQV-----
 410 420 430 440 450

gi-50913506. 480 490 500 510 520 530
 YSIDSRPTGNDVKEAYIESTKALTDTGTFSKSNEGSGTVLLETDIPNTKLGELPSTGSIG
 ||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |:
 gi-21909640. -IDIEEKLPD-EQGHSGSTTEIEDSKSSDVIIIGGQGEVVDTTE--DTQSGMTGHSGST-
 460 470 480 490 500

gi-50913506. 540 550
 TYLFKAIGSAAMIGAIGIYIVKRRKA
 | :| :| :| :|:
 gi-21909640. TEIEDSKSSDVIIIGGQGEVVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQQQIVETT
 510 520 530 540 550 560

gi-50913506.pep
 /home/morama/gas/pili/align/gi-28810263.pep

gi|28810263|dbj|BAC63201.1| protein F2-like protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 56 Initn: 82 Opt: 134 z-score: 74.7 E(): 1.3
 >>/home/morama/gas/pili/align/gi-28810263.pep (733 aa)
 initn: 82 init1: 56 opt: 134 z-score: 74.7 expect(): 1.3
 Smith-Waterman score: 155; 23.6% identity in 347 aa overlap
 (232-547:316-638)

gi-50913506. 210 220 230 240 250
 KITSLGSVLSYSLTELEPSYTKEAVNKTIVVSDNMSEGLITFNENSLIVEWKGKMAN
 ||| |::| |::| |::| |::| |:
 gi-28810263. EKEVAVDPAELKQINSECQQEITSVITWNQEVLTDE-KGMAYIVSVKEVDKNCELLEPKDY
 290 300 310 320 330 340

gi-50913506. 260 270 280 290 300 310
 ITEDGSVMVENTKIGIAKEVNNGFNLIFIY--DSLESISPNTSYKAVVNNKAIVGEE
 | ::| ::| |::| |::| |::| |::| |::| |::| |::| |::| |:
 gi-28810263. IKKEDGLTVNTNYV--KPTSGHYDIEVTFGNGHIDITEDTPDVI-VSGENQMKQIEGED-
 350 360 370 380 390

gi-50913506. 320 330 340 350
 GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKDP-DKGNGITSKEDSKIVYTYQ
 ::| | || ::| |::| |::| |::| |::| |::| |::| |::| |::| |:
 gi-28810263. SKPIDEVTEENNLLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSEQQQSGDMTIE
 400 410 420 430 440 450

360 370 380 390 400 410

FIGURE 58E

PCT/2005/27239

gi-50913506. IAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE

|| | | : | | : | : | : | | : | : | : | : |

gi-28810263. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE
460 470 480 490 500 510

420 430 440 450 460 470

gi-50913506. LKAPKGYSLNTE-TYEITAN-WTATVKTTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF

|| | | : | : | : | : | : | : | : | : | : |

gi-28810263. TAAPDGYEVATAITFTVNEQQVTVNGKATKGDAHIVMV-DAYKPTKGSGQV-----
520 530 540 550 560

480 490 500 510 520 530

gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNTKLGELPSTGSIG

|| : | : | : | : | : | : | : | : | : | : |

gi-28810263. --IDIEEKLKD-EQGHSGSTTEIEDSKSSDVIICGGQGEVVDTTE--DTQSGMTGHSGST-
570 580 590 600 610 620

540 550

gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA

| : | : | : |

gi-28810263. TKIEDSKSSDVIVGGQQIVETTEDTQTMHGDSGRKTEVEDTKLVQSFHFIDNKEPESN
630 640 650 660 670 680

gi-50913506.pep

/home/morama/gas/pili/align/orf80.pep

TRANSLATE of: orf80.seq check: 9824 from: 1 to: 1056

generated symbols 1 to: 352.

GETSEQ from morama, September 13, 2004 17:11.

SCORES Init1: 45 Initn: 69 Opt: 113 z-score: 70.8 expect(): 2.1

>>/home/morama/gas/pili/align/orf80.pep (352 aa)

initn: 69 init1: 45 opt: 113 z-score: 70.8 expect(): 2.1

Smith-Waterman score: 123; 22.8% identity in 311 aa overlap
(284-556:49-352)

260 270 280 290 300 310

gi-50913506. KMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNIKYAVVN-NKAIVGEE

:: | : | : | : | : | : | : | : |

orf80.pep ATALGTASLNQNVAETAGVVTGKSLQVTKMTYDDEVLMPETAFTFTIEDDMTASCKE
20 30 40 50 60 70

320 330 340 350 360 370

gi-50913506. GNPN-KAEFFYSMNPTKGNTYDNLDKKPDKGNGTTSKEDSKIVVYTYQIAFRKVDSVSKTP

| : | : | : | : | : | : | : | : | : | : |

orf80.pep GSLDIKNGLVEGLDKQVTVKYKNTDKTSQKTK-TAQFDFSKVKEPAIGVYRYMVSEKNDK
80 90 100 110 120 130

380 390 400 410 420

gi-50913506. LIGAIFGVYDTSNKLIDIVTTNKNG-----YAISTQ-VSSGKYKIKELKAPKGYSLN

| : | : | : | : | : | : | : | : | : | : |

orf80.pep KDGI---TYDDKKWTVDVYVGNKANNEEGFEVLYIVSKEGTSSTKPIEFTNSIKTTSIK
140 150 160 170 180 190

430 440 450 460

gi-50913506. TETYEITANW-----VTATVKTTSANSKSTTYTSDKNKATDNSEQVG----WLKNGI

| : | : | : | : | : | : | : | : | : |

orf80.pep IEK-QITGNAGDRKKSFNFTLTLQPSEYYKTGSVVKIEQDGSKKDTIGTPYKFTLGHGK
200 210 220 230 240 250

470 480 490 500 510

FIGURE 58F

PCT/US2005/02339

```

gi-50913506. FYSIDSRPTGNDV-----KEAYI-----ESTKALTDGTTFSKSNEGSGTVLLETDI
          ::| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
orf80.pep    SVMLSKLPIGINYYLSEDEANKDGYTTATLKEQGKEKSSDFTLSTQNQKTDESADIEIVV
          260      270      280      290      300      310

```

```

gi-50913506. PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRKA
| : :: || :|| . |: | : ||:: ||:: ||:: |
orf80.pep TNKRDTQVP-TGVVGT LAPFAVL SIVAIGGV-IYITKRKA
320          330          340          350

```

gi-50913506.pep

/home/morama/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492_5 EftLSL.A [Streptococcus pyogenes]

```
SCORES Init1: 45 Initn: 69 Opt: 109 z-score: 69.8 E(): 2.4  
>>/home/morama/gas/pili/align/gi-19224137.pep (342 aa)  
initn: 69 init1: 45 opt: 109 Z-score: 69.8 expect(): 2.4  
Smith-Waterman score: 169; 26.0% identity in 334 aa overlap  
(257-556:25-342)
```

	230	240	250	260	270	280
gi-50913506.	NKTVYVSDNMSEG	LTNFN	SLTVEWKGKMAN	ITEDGSVMVENTKIGIAKEVNNGFNL	SFI	
gi-19224137.	MKKNKL	LLATA	LATALGTASL	NQNVKAETAGVVSSGQLTI	KKSI	TN-FN---
	10	20	30	40		

290	300	310	320	330	340	
gi-50913506.	YDSLESI SPNISYKAVVN-NKAIVGEEGN-PNKAEFFYSNNP TKGNTYDNLDDKKPDKGNG					
gi-19224137.	DTTL--LMPKTDYTFSVNP DSAATGTESNLPIKPGIAVNQDIK-VSYSNTDKTSGKEKQ					
	50	60	70	80	90	100

350	360	370	380	390	400	
gi-50913506.	ITSKEDSKIVYTYQIAFRKVDSVSKTPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQV					
gi-19224137.	::: :: ::: : : : : : : : : : : : : :					
	VVV-DFMKVTFPSVGIIYRYVVTEK---	GTAEGVYDDTKWLVDVYVGNNIEKGGLEPKY				
	110	120	130	140	150	160

410 420 430 440 450
g1-50913506. SSGKYKIKELKAPKGY--SLNTETYELTANWVITATVKTTSANSKSTTYTSDKNKATDNS--
g1-19224137. IVSKKGDSATKEPIQFINSFETTSLKIEKE-VTGNTGDHKKAFTTLTLOPNEYYEAASSV

460 470 480 490
gi-50913506. --EQVGWLKN--GIFYST--DSR-----PTGND--VKEAYIE---STKALTDG
| : | : | : | : | : | : | : | : | : | : | : |
gi-19224137. VKIEENGQTKDVKIGEAYKFTLNDQSOSVILSKLPVGINYKVEEAANQGGYTTATLKDG
230 240 250 260 270 280

gi-50913506. RRKA

FIGURE 58C

gi-50913506.pep

/home/morama/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492_9 protein F2 [Streptococcus pyogenes]

SCORES Init1: 73 Initn: 73 Opt: 118 z-score: 68.9 E(): 2.7
 >>/home/morama/gas/pili/align/gi-19224141.pep (1161 aa)
 initn: 73 init1: 73 opt: 118 Z-score: 68.9 expect(): 2.7
 Smith-Waterman score: 174; 23.9% identity in 406 aa overlap
 (115-483:277-645)

90 100 110 120 130 140
 gi-50913506. VSLTETGPTSQEITTIANGINTGKIKPFSLENVSIISNGTATYNARGASVY--IALLTGAT
 ||:|||| :| :: : :|| |
 gi-19224141. IYTFTDYIAGLDKVQLSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTVNVLYGNE
 250 260 270 280 290 300

150 160 170 180 190
 gi-50913506. DGRTYNPILLAASYNGEGLNLVTKNIDSNSKSNLYGQTSVAKSSLPSITKKVTG-----T
 : : | | : | | |:: | : | : | : | : | : | : | : | : | : |
 gi-19224141. STKESNYITNGLSNVG-GSIESYNTETGEFVWYVYVNPNRNTNIPYATMNLWGGRARSNT
 310 320 330 340 350 360

200 210 220 230 240 250
 gi-50913506. ID---DVNKTTSLGSVLSYSITF--ELPSYTKEAVNKTIVVSDNMSEGLTFNFNSLTVE
 | |:| :: || : | : | : || | : | : | : | : | : | : | : |
 gi-19224141. SDLENDANTSSAELGEIQQVYEVPEGEKLPSSYGVDTKLTLRTD-TTAGLGNQFO----
 370 380 390 400 410

260 270 280 290 300 310
 gi-50913506. WKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNIKYAVVNNKAIVG
 |:: | : | : | | : | : | : | : | : | : | : | : | : | : |
 gi-19224141. -----MTKRQRIDFG--NNIQNKAFTIKV--TGKTDQSGKPLVVQSNLAS
 420 430 440 450 460

320 330 340 350 360
 gi-50913506. EEGNPNAEFFYSNNPTKGNTY--DNLDKPKDGNIGITSKEDSKIVYTY-----QTAF
 : | : | | : | : | : | : | : | : | : | : | : | : | : | : |
 gi-19224141. FRGASEYAAF---TPVGGNVYFQNEIALSPSKGSGSKSEFTKPSITVANILKRAQDRF
 470 480 490 500 510

370 380 390 400 410 420
 gi-50913506. RKVDSVSKTPLIGAIFGVYDTNSKLIDI-VTTNKNGYATSTQVSSGKYKIKELKAPGYS
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 gi-19224141. KKM-STDNVPLPEAAFLRSSNGNSQKLEASSNTQGEVHFKDLSGTYDLYETKAPGYQ
 520 530 540 550 560 570

430 440 450 460
 gi-50913506. -----LNTETYEIT-----ANWVT--ATVKTTSANSKSTTYTSDKNKATDNSEQVGWLKN
 | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 gi-19224141. QVTEKLATVTVDTTKPAEEMVTWGSFPHSSVKVEAN-KEVTIVNHETLTFSGKKI-WEND
 580 590 600 610 620 630

470 480 490 500 510 520
 gi-50913506. GIFYSIDSRPTGNDVKEAYIESTKALTDTTFSKSNEGSGTVLLETDPNTKLGELPSTG
 : | : | : | : |
 gi-19224141. ---RPDQRPARIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKQEYKYSVEE
 640 650 660 670 680

gi-50913506.pep

/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.1 E(): 3.8
 >>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)
 initn: 98 init1: 45 opt: 96 Z-score: 66.1 expect(): 3.8
 Smith-Waterman score: 181; 25.3% identity in 312 aa overlap
 (298-556:44-344)

270	280	290	300	310	320
gi-50913506.	NTKIGIAKEVNNGFNL	SFIYDSLESISPN	SYKAVVNNKAIVGEEGNPNKAEFFYSNNPT		
	20	30	40	50	60
gi-21909636.	ATALGTASLNQNVKAETAGVSENAKLIVKKT	FDSYTDNEVLMPKADYTFKVE	--ADSTA		
	80	90	100	110	120

330	340	350	360		
gi-50913506.	KGNTYDNL	DKKPDKGNGIT	SKEDSKIVYTYQIAFRKV		DSVSK
	130	140	150	160	170
gi-21909636.	SGTKDGLEIKPGIVNGLTEQIISYTNTDKPDSKV	KST-EFDFSKVVFPGIGVYRYTVSE			
	80	90	100	110	120

370	380	390	400	410	420
gi-50913506.	TPLIGAIFGV-YDTSNKLIDIVTTNKNGY	AI	STQVSSGKYKIKELKAPGY	--SLNTETY	
	130	140	150	160	170
gi-21909636.	KQ-GDVEGITYDTKKWTV	DVYVG	NKEGGFEPKFIVSKEQGTDVKKPVNFNNNSFATT	SL	
	80	90	100	110	120

430	440	450	460		
gi-50913506.	EITANWVTATVKTSAN	SKSTTYTSDK	NKATDNSE	QVGW	--LKNGI
	190	200	210	220	230
gi-21909636.	KVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVS	LQKGNEKFEV	KIGTPYKF	KLKNGE	
	80	90	100	110	120

470	480	490	500	510	
gi-50913506.	FYSIDSRTGNDVKEAYIESTKALT	DGTTFSKS	-NEGSGTVLL	--ETD	
	190	200	210	220	230
gi-21909636.	SIQLDKLPVGITYKV	NEMEANK	-DGYKTTASLKEGDGQSKMYQLDMEQKT	DESADEIV	
	80	90	100	110	120

520	530	540	550		
gi-50913506.	IPNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIV	VRRKA			
	190	200	210	220	230
gi-21909636.	VTNKRDQTQVP-TGVVGT	LAPFAVLSIVAIGGV	IYITKRKKA		
	80	90	100	110	120

gi-50913506.pep

/home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.0 E(): 3.8
 >>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)
 initn: 98 init1: 45 opt: 96 Z-score: 66.0 expect(): 3.8
 Smith-Waterman score: 181; 25.3% identity in 312 aa overlap
 (298-556:49-349)

PCT/US05/227239

	270	280	290	300	310	320
gi-50913506.	NTKIGIAKEVNNGFNLFSIYDSLES-ISPNIKYAVVNNKAIVGEEGNPKAEFFYSNNPT					
gi-28810259.	ATALGTASLNQNVKAETAGVSENAKLIVKKTFSYTDNEVLMPKADYTFKVE--ADSTA	20	30	40	50	60
						70
	330	340		350	360	
gi-50913506.	KGNTYDNLDKKPDKGNGIT-----SKEDSKIVYTYQIAFRKV-----DSVSK	80	90	100	110	120
gi-28810259.	SGKTKDGLEIKPGIVNGLTEQIISYTNTDKPSDKVKST-EFDFSKVVFPGLGVYRYTVSE					130
	370	380	390	400	410	420
gi-50913506.	TPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGY--SLNTETY	140	150	160	170	180
gi-28810259.	KQ--GDVEGITYDTKKWTVDVYVGNGKEGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSL					190
	430	440		450	460	
gi-50913506.	EITANWVTATVKTSAN-----SKSTTYTSDK---NKATDNSE-QVGW-----LKNGI	200	210	220	230	240
gi-28810259.	KVKKNVSGNTGELOKEFDFTLTNESTNFKDQIVSLQKGNEKFPEVKIGTPYKFKLKNGE					250
	470	480	490	500	510	
gi-50913506.	FYSIDSRTGNDVKEAYIESTKALTDTGTTESKS-NEGSGTVLL-----ETD-----	260	270	280	290	300
gi-28810259.	SIQLDKLPVGITYKVNEMEANK--DGYKTTASLKEGDGQSKMYQLDMEQKTDEADEIV					
	520	530	540	550		
gi-50913506.	IPNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA	310	320	330	340	
gi-28810259.	VTNKRDTQVP-TGUVGTLAPFAVLSIVAIGGV-IYITKRKKA					
gi-50913506.pep	/home/morama/gas/pili/align/gas15.pep					
GAS15 GAS15						
SCORES Init1: 42 Initn: 68 Opt: 96 z-score: 63.8 E(): 5						
>>/home/morama/gas/pili/align/gas15.pep (762 aa)						
initn: 68 init1: 42 opt: 96 z-score: 63.8 expect(): 5						
Smith-Waterman score: 96; 23.4% identity in 269 aa overlap (283-535:222-470)						
	260	270	280	290	300	
gi-50913506.	GKMANITEDGSVMVENTKIGIAKEVNNGFNLFSIYDSLES-ISPNIKYAVVNNKAIVGEEGNPKAEFFYSNNPTKGN	200	210	220	230	240
gas15.pep	VWYYSDNAPISNPDESFKRESESNLVSTSQSLMRQALKQLIDPNLATKMPKQVPDDFOL					250
	310	320	330	340	350	360
gi-50913506.	AIVGEEGNPKAEFFYSNNPTKGN	260	270	280	290	300
gas15.pep	YDNLDKKPDKGNGITSKEDSKIVYTYQIAFRKV--V SIFESEDKGDKYNKGYQNLLSGGLVPT--KPPTPGDPPMPPNQPO--TTSVLIRKYAT					
	370	380	390	400	410	420

FIGURE 58J

PCT Zusatz 222 235

```

gi-50913506. DSVSKTPLIGAIFGVY-DTSNKL-IDIVTTNKNGYAISTQVSSGKYKIKELKAPKGYSLN
          : ||| | ||| : : | : ||| : : ||| | | | : ||| : ||| : ||| : ||| : ||| : |
gas15.pep   GDYSKL-LEGATLQLTGDNVNSFQARVFSSNDIGERI--ELSDGTYTLTELNSPAGYSIA
          310      320      330      340      350      360

```

```

gi-50913506. TE-TYEITANWVTATV--KTSANSKSTTYTSDKNKATDNSEQVGWLKN---GIFYSIDSR
| : :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
gas15.pep EPITFKVEAGKVYTIDGKQIENPNKEIVEPYVSVEAYNDFEEFSVLTQNYAKFYYYAKNK
370      380      390      400      410      420

```

gi-50913506.	480	490	500	510	520	530
	PTGNDVKEAYIESTKALT	DGTTFSKSNEGSGTVLLET	DIPNTKL	GELPSTGSIGTYLFKA		
gas15.pep	:::: : : :	:: : :	: :			
	NGSSQVVYCFNADLKSPD	-----	SEDDGGKTMT	-----	PDFTTGEVKYTHIAGRDLFKY	
	430	440	450	460	470	

540 550

qi-50913506. IGSAAAGATGATGTTVKRRKA

gas15.pep TVKPRDTDPTFLKHIIKKVIEKGYREKGQAIYEYSGLTETQLRAATQLATYYFTDSAELDK
480 490 500 510 520 530

gi-50913506.pep

/home/morama/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

```

SCORES Init1: 41 Initn: 41 Opt: 87 z-score: 63.6 E(): 5.2
>>/home/morama/gas/pili/align/gi-13621428.pep (340 aa)
initn: 41 init1: 41 opt: 87 Z-score: 63.6 expect(): 5.2
Smith-Waterman score: 109; 22.6% identity in 345 aa overlap
(256-556:17-340)

```

230 240 250 260 270 280
gi-50913506. VNKTIVVSDNMSEGLTNFNSLTVEWKGMANITEDGSVMVENTKIGIAKEVNNGFNLSF
| | | : : | : : | : ||:
gi-13621428. MKLRHLLLTTGAALTSEFAATTVHGTVVNGAKLTVTK-----NLDD

290. 300. 310. 320. 330. 340.
gi|50913506. IYDSELSISPNTSYKAVVNNKAIVGEEGNPNAEFFYSNNPTKGNTYDNLDKKPDKGNGL
gi|13621428. VNSN- ALIPPTDFTKIEEDITVNEEGNKFRGVAL- NTPMTKVTYTNSDK- GGSNTIK

350 360 370
gi-50913506. TSKED-SKI-----VYTYQTAFRKVDSV-----SKTPLIGAIF
| : | | : | | : | | : | | : | | : | | : | | : | |
gi-13621428. TAEFDSEVTFEKPGVYYYYKVTEEKIDKVPGVSYDTSYTVQHVWLNEEQKPVATYIV

gi-50913506. GYVDTSNKLIDIVTT-NKNGYAISTQVS-SGKYKIKELKAPKGYSINTETYEITANWTA
| : | : | :: :: : :: : || : | : | :: : | : | :: : | : | :: : |
gi-13621428. GYKEGSKVPIQFKNSLDSTTLTVKKVSGTGGDRSKDFNE--GLTLKANOYYKASEKVM

440 450 460 470 480 490
gi-50913506. TVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIFYSIDSRPTGND--VKEAYIESTKAL

FIGURE 58V

PCT/US05/23230
gi-13621428. E-KTTKGQAPVQT---EASIDQLYHFTLKDGESIKVTNLPGVVDYVVTEDDYKSEKYT
220 230 240 250 260

500 510 520 530 540
gi-50913506. T-----DGT-----FSKSNEGSGTVILLETDIPNTKLGELPSTGSIGTYLFKAIGSAA
| |::| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-13621428. TNVEVSPQDGAVKNIAGNSTEQETSTDKDMTITFTNKKDFEVPTGVAMTVAPYIALGIVA
270 280 290 300 310 320

550
gi-50913506. MIGAIGIYIVKRRKA
: | :| :| :| :|
gi-13621428. VGGAL--YFVKKKNA
330 340

gi-50913506.pep
/home/morama/gas/pili/align/gi-19224135.pep

gi|19224135|gb|AAL86406.1|AF447492_3 Cpa [Streptococcus pyogenes]

SCORES Init1: 41 Initn: 41 Opt: 86 z-score: 61.0 E(): 7
>>/home/morama/gas/pili/align/gi-19224135.pep (756 aa)
initn: 41 init1: 41 opt: 86 z-score: 61.0 expect(): 7
Smith-Waterman score: 101; 19.6% identity in 306 aa overlap
(243-535:193-462)

220 230 240 250 260 270
gi-50913506. SLTFELPSYTKEAVNKTVYVSDNMSEGLTFNFNSLTVEWKGKMANITEDGSVMVENTKIG
: : : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-19224135. PKNANGYMDKIEPLNAILVTQQAVWYYSDDSYGNIKTLWASEL---KDGKIDFEQVKL-
170 180 190 200 210
280 290 300 310 320 330
gi-50913506. IAKEVNNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPN--KAEFFYSNNPTKGN
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-19224135. -MREAYS---KLISDDLEETSKNKLPQGSKLNIFVPQDKSVQNLLSAEYVPESPAPGQ
220 230 240 250 260 270
340 350 360 370 380
gi-50913506. TYDNLDKPKDKNGNITSKEDSKIVYTY-QIAFRKVDSVSKTPLIGAIFGVYDTSNKLIDI
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-19224135. S---PEPP---VQTKKTSVITRKYAEFDYSKILLEGATLRLTGE-DILDFQEKG---V
280 290 300 310
390 400 410 420 430 440
gi-50913506. VTTTNKNGYAIISTQVSSGKYKIKELKAPKGYSINTET-YEITANWTATVKTTSANSKSTT-
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-19224135. FQSNGTGEGKI--ELSNGTYTLTETSSPDGYKIAEPIKFRVVNKKFIVQKDGSOVENPNK
320 330 340 350 360 370
450 460 470 480 490
gi-50913506. ---YTSOKNATDNSEQVG---WLKNGIFYSIDSRPTGNDVKEAYTESTKALTIDGTTF
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-19224135. EVAEPYSVEAYSDMQDSNYINPETFTPYGFYYAKNKKDKSSQVYCFN-----ADLHSP
380 390 400 410 420 430
500 510 520 530 540 550
gi-50913506. SKSNEGSGTVILLETDIPNTKLGELPSTGSIGTYLFKAIGSAA MIGAIGIYIVKRRKA
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-19224135. PESEDGGGTI--DPDISTMK--EVKYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKG
440 450 460 470 480

FIGURE 58L

gi-19224135. YNKKGDSYNGLTETQFRAATQLATYYFTDSTDLKTLKTYNNGKGYHGFESMDEKTLAVTK
 490 500 510 520 530 540

gi-50913506.pep
 /home/morama/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

SCORES Init1: 70 Initn: 149 Opt: 86 z-score: 60.1 E(): 7.8
 >/home/morama/gas/pili/align/gi-50913505.pep (1036 aa)
 initn: 149 init1: 70 opt: 86 Z-score: 60.1 expect(): 7.8
 Smith-Waterman score: 120; 21.5% identity in 469 aa overlap
 (33-428:503-966)

	10	20	30	40	50	60
gi-50913506.	NRRETVREKILITAKKMLACLAIALAVVGLGMTRVS-ALKDDTAQLKITNIEGGPTVTL					
	480	490	500	510	520	
gi-50913505.	YSQWVYHGTFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDAVKNSLLGVNG----L					

	70	80	90	100	110	
gi-50913506.	YKIGEGVYNTNGDSFINFK---YAEGVSLTETGPTSQEIT-TIANGINTGKIKPFSTEN					
gi-50913505.	LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNS					
	530	540	550	560	570	580

	120	130	140	150	160	
gi-50913506.	VSISNGTATYNARGASVYIALLTGAT-----DGRTYNPILLAAS-YNGEGNLVTK					
gi-50913505.	LLDPNTLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPFYFGEDGYRSG					
	590	600	610	620	630	640

	170	180	190	200	210	
gi-50913506.	NIDS--KSNEYLYGQ--TSVA-KSSLPSITKKVTGTTDDVNKKTTSLGSVLSYSLTFE					
gi-50913505.	NGSSNDRNNVTRSQEWSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSE					
	650	660	670	680	690	700

	220	230	240	250	260	
gi-50913506.	LPSY-TKEAVNKIVY-VSDNMSEGGLTF-NFNSLTVEWKGRMANITE					
gi-50913505.	EHYYGITDTAELEKTLNKIVEDSKSQLGISDSLSQLVYDYYDKQPDVLVTRSKVNDETE					
	710	720	730	740	750	760

	270	280	290			
gi-50913506.	DGSVMVENTKIGIAKEVNNGFNLNSFIYDS-LESIISPNT-					
gi-50913505.	ILYQKDQVQEAGKDIIDKVVFPTKTTSQPKGVTLTFKSDYKVDETYTLSFNVKASDE					
	770	780	790	800	810	820

	300	310	320	330	340	350
gi-50913506.	SYKAVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDDKKPDK-GNGITSKEDSK-					
gi-50913505.	AYEKYKDNEGRYSEMGSDDTDYGTNQTSGGKGLPSNSDASVNYMADGREQKLKYHPVI					
	830	840	850	860	870	880

	360	370	380	390	400	
gi-50913506.	IVYTYQTAFRKVDS-VSKTPLIGAIFGVYDTSNKLI-DIVTTNKNGYAIISTQVSSGK-YK					

FIGURE 58M

PCT/US05/27239

gi-50913505. QVKTVPITFTKVDADNNQKKLAGVEFELRKEDKKIVWEKTTGSNGQLNFKYLQKGKTYY
 890 900 910 920 930 940

gi-50913506. IKELKAPKGYSLNTEETYEITANWVTATVTSANSKSTTYTSDKNKATDNSEQVGWLNGI
 : | || ||:: : :::::
 gi-50913505. LYETKAALKGYTLPEPNPWEVAVANNGDIKVKHPIEGELSKDGSYMIRKNYKTYQLPSSGR
 950 960 970 980 990 1000

gi-50913506.pep

/home/morama/gas/pili/align/gi-13621430.pep

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 43 Initn: 67 Opt: 67 z-score: 59.2 E(): 8.7
 >>/home/morama/gas/pili/align/gi-13621430.pep (215 aa)
 initn: 67 init1: 43 opt: 67 z-score: 59.2 expect(): 8.7
 Smith-Waterman score: 67; 27.8% identity in 90 aa overlap
 (433-522:60-143)

410 420 430 440 450 460
 gi-50913506. VSSGKYKIKELKAPKGYSLNTEETYEITANWVTATVTSANSKSTTYTSDKNKATDNSEQV
 : | || ||:: : :::::
 gi-13621430. TASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLPNSVT-TSVKGNGKTSFEQL
 30 40 50 60 70 80

470 480 490 500 510 520
 gi-50913506. GWLKNGLIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSTVLLIEDIPNTKLGE
 : : | :: : | :: : | :: : : | ||: ||||: |||:: :: ||||
 gi-13621430. TFSEVGQYHYKIHQLLGKNSQYHYDETVEVVIYVLY---NEQSGA--LETNLVSNKLGE
 90 100 110 120 130 140

530 540 550
 gi-50913506. LPSTGSIGTYLFKAIGSAAMTGAIGIYIVKRRKA

gi-13621430. TEKSELIFKQEYSEKTPEPHQPDTEKEKPKQRNGILPSTGEMVSÝVSALGIVLVATIT
 150 160 170 180 190 200

gi-50913506.pep

/home/morama/gas/pili/align/gi-19745303.pep

gi|19745303|ref|NP_606439.1| hypothetical protein [Streptococcus pyogenes MGAS82
 321]

SCORES Init1: 51 Initn: 106 Opt: 69 z-score: 58.4 E(): 9.5
 >>/home/morama/gas/pili/align/gi-19745303.pep (344 aa)
 initn: 106 init1: 51 opt: 69 z-score: 58.4 expect(): 9.5
 Smith-Waterman score: 129; 24.0% identity in 308 aa overlap
 (298-556:44-344)

270 280 290 300 310 320
 gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
 : : : :: : | :: : : | :| : :: :
 gi-19745303. ATALGTASLNQNVKAETAGVIDGSTLVVKTFPSYTDDKVLMPKADYTFKVE---ADDNA
 20 30 40 50 60 70

330 340 350 360 370
 gi-50913506. KGN TYDNL DKK PDK GNGI-----TSKED SKI VTY QIA FRK VD SV SKT -- P
 ||:| |:|| || :|:| :|| | :| | |

201/487

gi-197433037 KGKTKDGLDKPGVLDGLENTKTIHGNSDKTTAKEKSVNFDANVKFPGVGVYRTVSE
80 90 100 110 120

80 90 100 110 120 130

380 390 400 410 420
gi-50913506. LIGAIFGV-YDTSNKLIDIVTTNKN--GYAISTQVSS-GKYKIKELKAPKGYSINTETYE

gi-19745303. VNGNKAGIAYDSQQWTVDVYVNVNREDGGFEAKYIVSTEGGQSDKKPVLFKNF-FDTTSLK
140 150 160 170 180

430 440 450 460 470
gi-50913506. ITANWVTATVKTSANSKTTYTSDKNKATDNSEOVGWLKNGIF-----YSIDSP-----

gi-19745303 VTKK-VTGNTGEHQRSFSFTLLLTPNECFEKQGVVNILQGGETKKVVIGEYSFTLKDKE
190 200 210 220 230

480 490 500 510

gi-197453-03. SVTLSQLPGIEYKVTEEDVTKGYKTSATLKDGFVTDGYNLGDKSTTDESMDELYTWW

520 530 540 550

gi-50913506. KLGELPSTGSIGTYLFKAIGSAAMIGATGITYIVKRRKA

gi-19745303. RDTQVP-TGVVGTTLAPPFAVLSIVAIGGV-IYITKRKKA

340

! Distributed over 1 thread

Start time: Wed Sep 15 18:45:54 2004
Completion time: Wed Sep 15 18:45:54 2004

! Completion time: Wed Sep 15 18:46:02 2004

! Database scan: 0:00:00.1
! Post-scan processing: 0:00:01.9

Total CPU time: 0:00:02.0
Output File: gi-50913506.fasta

: Output File: gi-50913506.fasta

FIGURE 580

! !SEQUENCE_LIST 1.0505 / 27235

(Peptide) FASTA of: gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences

z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	

< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: * *
40	0	4: * *
42	0	5: * *
44	1	5:= * *
46	7	5:====*==
48	21	5:====*=====
50	7	5:====*==
52	2	4:== * *
54	4	3:==*=
56	3	3:==*
58	0	2: *
60	4	2:==*==
62	1	2:==*
64	0	1:*
66	0	1:*
68	0	1:*
70	0	1:*
72	0	0:
74	0	0:
76	0	0:
78	0	0:
80	3	0:====
82	2	0:==
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:
100	0	0:

FIGURE 59

102 P00T/LIG05/27239
 104 0 0:
 106 0 0:
 108 0 0:
 110 0 0:
 112 0 0:
 114 0 0:
 116 0 0:
 118 0 0:
 >120 1 0:=

Joining threshold: 36, opt. threshold: 24, opt. width: 16, reg.-scaled

The best scores are:

	init1	initn	opt	z-sc E(55)..
/home/morama/gas/pili/align/gi-13621430.pep	Begin: 1	End: 215		
! gi 13621430 gb AAK33240.1 hypothetical protein [Streptococcus pyogenes]	1338	1338	233.9	1.8e-09
/home/morama/gas/pili/align/gi-19745305.pep	Begin: 1	End: 193		
! gi 19745305 ref NP_606441.1 hypothetical protein [Streptococcus pyogenes]	163	243	273	82.2
/home/morama/gas/pili/align/gi-28810261.pep	Begin: 2	End: 187		
! gi 28810261 dbj BAC63199.1 hypothetical protein [Streptococcus pyogenes]	164	239	268	81.5
/home/morama/gas/pili/align/gi-19224139.pep	Begin: 2	End: 187		
! gi 19224139 gb AAL86410.1 AF447492...	164	236	265	81.0
/home/morama/gas/pili/align/orf82.pep	Begin: 30	End: 222		
! TRANSLATE of: orf82:seq check: 4296...	163	235	264	81.0
/home/morama/gas/pili/align/gi-21909638.pep	Begin: 2	End: 181		
! gi 21909638 ref NP_663906.1 hypothetical protein [Streptococcus pyogenes]	164	239	261	80.5
/home/morama/gas/pili/align/gi-19745303.pep	Begin: 84	End: 183		
! gi 19745303 ref NP_606439.1 hypothetical protein [Streptococcus pyogenes]	121	121	126	61.4
/home/morama/gas/pili/align/gi-13621428.pep	Begin: 6	End: 174		
! gi 13621428 gb AAK33238.1 hypothetical protein [Streptococcus pyogenes]	58	86	122	60.9
/home/morama/gas/pili/align/gi-19224137.pep	Begin: 93	End: 201		
! gi 19224137 gb AAL86408.1 AF447492...	88	88	119	60.4
/home/morama/gas/pili/align/gi-50913503.pep	Begin: 549	End: 625		
! gi 50913503 ref YP_059475.1 Fibronectin type III domain containing protein 1	73	73	117	60.4
/home/morama/gas/pili/align/gi-19224134.pep	Begin: 631	End: 697		
! gi 19224134 gb AAL86405.1 AF447492...	73	73	115	60.1
\End of List				7.8

gi-13621430.pep

/home/morama/gas/pili/align/gi-13621430.pep

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 1338 Initn: 1338 Opt: 1338 z-score: 233.9 E(): 1.8e-09
 >>/home/morama/gas/pili/align/gi-13621430.pep (215 aa)
 initn: 1338 init1: 1338 opt: 1338 z-score: 233.9 expect(): 1.8e-09
 Smith-Waterman score: 1338; 100.0% identity in 215 aa overlap
 (1-215:1-215).

	10	20	30	40	50	60
gi-13621430.	MKK	SILRILAIGYLLMSFC	LDSVEAENLTASINIEV	I	QVDVATNKQSSDIDET	TMFVI
gi-13621430.	MKK	SILRILAIGYLLMSFC	LDSVEAENLTASINIEV	I	QVDVATNKQSSDIDET	TMFVI
	10	20	30	40	50	60
	70	80	90	100	110	120
gi-13621430.	EALDKESPLPNSVTT	TKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDET	VYEVV			
gi-13621430.	EALDKESPLPNSVTT	TKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDET	VYEVV			

FIGURE 59A

PCT/LG020/278039

90

100

110

120

gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTEKEKPKOKKRNGI

gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTEKEKPQKKRNGI
130 140 150 160 170 180

190 200 210
qi-13621430 LPSTGEMVSYVGALCIVI VTTVY KCVHMM

gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK

gi-13621430.pep

/home/morama/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP_606441.1| hypothetical protein [Streptococcus pyogenes MGAS8232]

```

SCORES Init1: 163 Initn: 243 Opt: 273 z-score: 82.2 E(): 0.5
>>/home/morrama/gas/pili/align/gi-19745305.pep (195 aa)
initn: 243 init1: 163 opt: 273 Z-score: 82.2 expect(): 0.5
Smith-Waterman score: 320; 31.9% identity in 213 aa overlap
(1-213:1-193)

```

	10	20	30	40	50	60
gi-13621430.	MKK	SILRILAIGYLLMSFC	LDSVEAENLTASINIEV	INQVDVATNKQSSDIDET	FMFVI	
		:::: ::	: : :: :: :	:: :		:
gi-19745305.	M	RKYWKMLFSVVMILTMLAFN	QTVLAKDSTVQTSISVEN	VLERAGDSTS	-----	FSVAL
	10	20	30	40	50	

	70	80	90	100	110	120
gi-13621430.	EALDKESPLPNSVTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV :: : : : : :: : : :: : :: : :: :					
gi-19745305.	ESTIDAMKTI-DEIT--IAGSGKASFPLTFTTUVGQYTDRVYQKPSQNKDYQADTTVFDVL 60	70	80	90	100	110

190 200 210
gi-13621430 LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK

gi-19745305. LPLAGEVKSLLGILSIVLLGLLVLLYV-KKEKSRL
170 180 190

gi-13621430.pep

/home/morama/gas/pili/align/qi-28810261.pen

gi|28810261|dbj|BAC63199.1| hypothetical protein [Streptococcus pneumoniae GGI_1111]

```
SCORES Init1: 164 Initn: 239 Opt: 268 z-score: 81.5 E(): 0.55  
>>/home/morama/gas/pili/align/gi-28810261.pep (189 aa)  
initn: 239 init1: 164 opt: 268 Z-score: 81.5 expect(): 0.55  
Smith-Waterman score: 306; 30.6% identity in 206 aa overlap
```

FIGURE 59B

205/487

(8-2132-187) USGS / 27239

	10	20	30	40	50	60
gi-13621430.	MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETMFVI	:::: :: : : :: :: :: :: :: :: :: :: ::				
gi-28810261.	MLFSVVMLTMLAFNQTVLAKDSTVQTISIVENVLERAGDSTP-----FSIAL					
	10	20	30	40		
	70	80	90	100	110	120
gi-13621430.	EALDKESPLPNSVTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETIVYEVV	:: : : : : : : : : :; :: : : : :				
gi-28810261.	ESIDAMKTIEE---ITIAGSGKASFSPLTFTTVGQTYTYRVYQKPSQNKDYQADTTVFDV	50 60 70 80 90 100				
	130	140	150	160	170	180
gi-13621430.	IYVLYNEQSGALETNLVSNKLGTEKESELIFKQEYSEKTPPEPHQPDTEKEKPQKKRNGI	: :				
gi-28810261.	VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNI-----PKTP--	110 120 130 140 150				
	190	200	210			
gi-13621430.	LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK	: :				
gi-28810261.	LPLAGEVKSLLGILSTVLLGLVLLYYV-KKLKSRL	160 170 180				

gi-13621430.pep
/home/morama/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]

```
SCORES Init1: 164 Initn: 236 Opt: 265 z-score: 81.0 E(): 0.57  
>>/home/morama/gas/pili/align/gi-19224139.pep (189 aa)  
initn: 236 init1: 164 opt: 265 Z-score: 81.0 expect(): 0.57  
Smith-Waterman score: 303; 30.6% identity in 206 aa overlap  
(8-213:2-187)
```

10	20	30	40	50		
gi-13621430.	MKKSTILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSSDIDETFMFVI	:::: :: . : : :: :: :: :: :: :: :: ::				
gi-19224139.	MLFSVVAMILTMLAFNQTVLAKDSTVQTTSISVENVLERAGDSTP-----FSIAL					
	10	20	30	40		
	70	80	90	100	110	120
gi-13621430.	EALDKESPLPNCSVITTSVKGNNGKTISFEQLTFSEVGQYHYKIHOLLGKNSQYHYDETIVYEVV	:: :: :: :: :: :: :: :: :: :: :: :: :: ::				
gi-19224139.	ESIDAMKTEEE---TTAGSGKASFSPLTFTTVGQTYRVTYQKPSQNKDYQADTTVFDVL					
	50	60	70	80	90	100
	130	140	150	160	170	180
gi-13621430.	IVVLYNEQSGALETNLVSNLGETEKSELIFKQEYSEKTPPEPHQPDTEKEKPQKKRNGI	:: :: :: :: :: :: :: :: :: :: :: :: ::				
gi-19224139.	VVVTYDED-GTLVAKVISRRAGDEEKSAITFKPKRLVVKPIPPRQPNI-----PKTP--					
	110	120	130	140	150	
	190	200	210			
gi-13621430.	LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK	:: :: :: :: :: :: :: :: :: :: :: ::				
gi-19224139.	LPLAGEVKSLLGILSIVLLGLLVLLYV-KKLKSKL					
	160	170	180			

FIGURE 59C

PCT/US05/27239

gi-13621430.pep

/home/morama/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672
 generated symbols 1 to: 224.

GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 163 Initn: 235 Opt: 264 z-score: 81.0 E(): 0.58
 >>/home/morama/gas/pili/align/orf82.pep (224 aa)
 initn: 235 init1: 163 opt: 264 z-score: 81.0 expect(): 0.58
 Smith-Waterman score: 304; 30.5% identity in 213 aa overlap
 (1-213:30-222)

gi-13621430.	10 20 30
	MKKSILRILAIGYLLMSFCLLDSVEAENLTA :: ::::: :: :::: :: ::
orf82.pep	LLFQRVKIFLLTIVLSVLFKNNEERRRLLRKYWKMLFSVVMILITMLAFNQTVLAKDSTV 10 20 30 40 50 60
gi-13621430.	40 50 60 70 80 90
orf82.pep	SINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLPNSVTTSVKGNGKTSFEQLTFS :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: QTSISVENLERAGDSTP-----FSVALESIDAMKTI-DEIT--IAGSGKASFSPLTFT 70 80 90 100 110
gi-13621430.	100 110 120 130 140 150
orf82.pep	EVGQYHYKIHQLLGKNSQYHYDETIVYEVVIYVLYNEQSGALETNLVSNKLGTEKESELIF :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: TVGQTYRYVYQKPSQNQDYQADTTVFDVLVYVTYDED-GTLVAKVISRRAGDEEKSATF 120 130 140 150 160 170
gi-13621430.	160 170 180 190 200 210
orf82.pep	KQEYSEKTPEPHQPDTEKEKPKQRNGILPSTGEMVSYVSAIGIVLVATITLYSIYKKL : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: KPKRLVKPKIPPRQPNI-----PKTP---LPLAGEVKSLLGILSIVLLGLIVLLYV-KKL 180 190 200 210 220
gi-13621430.	KTSK
orf82.pep	: KSRL

gi-13621430.pep

/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP_663906.1| hypothetical protein [Streptococcus pyogenes MGAS31
 5]

SCORES Init1: 164 Initn: 239 Opt: 261 z-score: 80.5 E(): 0.62
 >>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)
 initn: 239 init1: 164 opt: 261 z-score: 80.5 expect(): 0.62
 Smith-Waterman score: 302; 31.5% identity in 200 aa overlap
 (14-213:2-181)

gi-13621430.	10 20 30 40 50 60
	MKKSILRILAIGYLLMSFCLLDSVEAENLTA :: ::::: :: ::: :: :: ::
gi-21909638.	MILTMLAFNQTVLAKDSTVQTSISVENLERAGDSTP-----FSIAL

207/487

PCT/US05/27219

20 30

40

70	80	90	100	110	120
gi-13621430.	EALDKESPLPNSVTTSVKGNKGTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEYEVV	: :: : : : : :: : :: : : :			
gi-21909638.	ESIDAMKTEE---ITIAGSGKASFSPLETTTVGQTYFVYQKPSQNQDYQADTTVFDVL				
	50	60	70	80	90

	130	140	150	160	170	180
gi-13621430.	LYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTEKEKPQKKRNGI	: : ; : : : : : : : : : :				
gi-21909638.	VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNI-----PKTP-----	100	110	120	130	140

	190	200	210	
gi-13621430.	LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK	: : :: : : : : : : :		
gi-21909638.	LPLAGEVKSLLGILSTVLLGLLVLLYV-KKLKSRL			
	150	160	170	180

gi-13621430.pep

/home/morama/gas/pili/align/gi-19745303.pep

gi|19745303|ref|NP_606439.1| hypothetical protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 121 Initn: 121 Opt: 126 z-score: 61.4 E(): 6.7
>/home/morama/gas/pili/align/gi-19745303.pep (344 aa)
initn: 121 init1: 121 opt: 126 Z-score: 61.4 expect(): 6.7
Smith-Waterman score: 126; 27.0% identity in 100 aa overlap
(59-155; 84-183)

gi-13621430, LTASTNIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLP--NSVTTSVKGNG-KTSF
 ||::|:: : : | | |::| : : : |
 gi-19745303: LMPKADYTFKVEADDNAKGKTKDGLDIKPVGVIDGLENTKTIHGNSDKTTAKEKSVNPDF
 60 70 80 90 100 110

90	100	110	120	130	140
gi-13621430.	EQLTFSEVGQYHYKIHQLLGKNSQYHYDETIVYEVVIYVLYNEQSGALETNLVSNKLGETE				
	::				
gi-19745303.	ANVKFPGVGVYRTVSEVNGNKAGIAYDSQQWTVDVYVNVNREDGGFEAKYLVSTEGGQSD				
	120	130	140	150	160

	150	160	170	180	190	200
gi-13621430.	KSELTFKQEYSEKTPEPHQPDTEKEKPQKKRNGILPSTGEMVSYVSALGIVLVATITLY	: :: : ::				
gi-19745303.	KKPVLFKNFFDTTSLKVTKVTGNTGEHQRSFSFTLLTPNECFEKQGVVNLIQGGETKK	180	190	200	210	220

gi-13621430.pep
/home/morama/qas/pili/alijm/gi-13621428.pep

qi|13621428|gb|AAK33238-11 hypothetical protein [Streptococcus pneumoniae]

```

SCORES Init1: 58 Initn: 86 Opt: 122 z-score: 60.9 E(): 7.2
>>/home/morama/gas/pili/align/gi-13621428.pep (340 aa)
initn: 86 init1: 58 opt: 122 Z-score: 60.9 expect(): 7.2
Smith-Waterman score: 135: 29.1% identity in 172 aa overlap

```

FIGURE 59E

(8-159E6#17D) / US05 / 27239

	10	20	30	40	50
gi-13621430.	MKKSTILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVD-VATNKQSSDIDETFMF-	: : : : : : : : : : : : :			
gi-13621428.	MKLRHILLTGAALTSFAAT-TVHGETVNGAKLTVKNLDLVNSNALIPNTDFTFKIE	10 20 30 40 50			

	60	70	80	90	100
gi-13621430.	VIEALDK-----ESPLPN-SVTTSVKGNGKTSFEQLTFSEV-----GQYHYKL-H	: :			
gi-13621428.	PDTITVNEGDGNKFKGVALNTPMTKVYTNTNSDKGGSNKTAEFDFSEVTFEKPGVYYYKVTE	60 70 80 90 100 110			

	110	120	130	140	150	160
gi-13621430.	QLLGKNSQYHYDETIVYEVVIVLYNE-QSGALETNLVSNLGETEKSELIFKQEYSEKTP	: : : : : : : : : : : : : : : : : : :				
gi-13621428.	EKIDKVPGVSYDFTTSYTVQVHLWNEEQQKPVATYIVGYKEGS--KVPIQFKNSLDSTTL	120 130 140 150 160 170				

	170	180	190	200	210
gi-13621430.	EPHQPDTEKEKPQKKRNGILPSTGEMVSYSVSLGIVLVATITLYSIYKKLKTSK	180 190 200 210 220 230			
gi-13621428.	TVKKVSGTGGDRSKDFNFGILTAKQYYKASEKVMILEKTTKGQAPVQTEASIDQLYHF				

gi-13621430.pep

/home/morrama/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492_5 EftLSL.A [Streptococcus pyogenes]

SCORES Init1: 88 InitN: 88 Opt: 119 z-score: 60.4 E(): 7.5
 >>/home/morrama/gas/pili/align/gi-19224137.pep (342 aa)
 initn: 88 init1: 88 opt: 119 Z-score: 60.4 expect(): 7.5
 Smith-Waterman score: 119; 29.7% identity in 111 aa overlap
 (72-176:93-201)

	50	60	70	80	90
gi-13621430.	DVATNKQSSDIDETFMFVIEALDKESPLPNVTTSVKGNGK-----TSFEQLTFSEVGQY	: :			
gi-19224137.	SVNPDSAATGTESNLPIKPGTAVNNQDIKVSVSNTDKTSGKEKQVVVDFMKVTFPSVGIY	70 80 90 100 110 120			

	100	110	120	130	140	150
gi-13621430.	HYKIHQLLGKNSQYHYDETIVYEVVIVLYNE-QSGALETNLVSNLGETEKSELIFKQEY	: :				
gi-19224137.	RYVVTEKGTAEGVTYDDTKWLVBVYVGNNEKGG-LEPKYIVSKKGDSATKEPIQFNNSF	130 140 150 160 170 180				

	160	170	180	190	200	210
gi-13621430.	SEKTPPEPHQPDTEKEKPQKKRNGILPSTGEMVSYSVSLGIVLVATITLYSIYKKLKTSK	: :				
gi-19224137.	-ETTSLKIEKEVTGNTGDHKKAFTFTLTLQPNEYYEASSVVKIEENGQTDVDKIGEAYKF	190 200 210 220 230 240				

gi-13621430.pep

/home/morrama/gas/pili/align/gi-50913503.pep

gi|50913503|ref|YP_059475.1| Fibronectin-binding protein [Streptococcus pyogenes MGAS10394]

PCT/US05/27239

SCORES Init1: 73 Initn: 73 Opt: 117 z-score: 60.4 E(): 7.6
 >>/home/morama/gas/pili/align/gi-50913503.pep (627 aa)
 initn: 73 init1: 73 opt: 117 z-score: 60.4 expect(): 7.6
 Smith-Waterman score: 118; 28.7% identity in 87 aa overlap
 (129-215:549-625)

100	110	120	130	140	150
gi-13621430.	KIHQLLGKNSQYHYDET	VYEVVIVLYNEQSGALET	NLVSNKLGETEKSELIFKQEYSEK		
			: : : :		
520	530	540	550	560	570
gi-50913503.	IETEDTKEPEVLMGGQSESVEFTKD	TQMSGFSETATVV	---EDTRPKLVFHFDNNEP		
160	170	180	190	200	210
gi-13621430.	TPEPHQPDTEKEKPQKKRNGILPSTGEMVS	YVSVSALGIVLVATITLYSIYKKLKTSK			
	: : : : : :	: : : : :			
580	590	600	610	620	
gi-50913503.	KVEEN	REKPTKNITPILPATGDIENVLAFLGILILSVLSIFSLLKNQSNKKV			

gi-13621430.pep

/home/morama/gas/pili/align/gi-19224134.pep

gi|19224134|gb|AAL86405.1|AF447492_2 protein F [Streptococcus pyogenes]

SCORES Init1: 73 Initn: 73 Opt: 115 z-score: 60.1 E(): 7.8
 >>/home/morama/gas/pili/align/gi-19224134.pep (698 aa)
 initn: 73 init1: 73 opt: 115 z-score: 60.1 expect(): 7.8
 Smith-Waterman score: 115; 27.4% identity in 73 aa overlap
 (143-215:631-697)

120	130	140	150	160	170
gi-13621430.	DET	VYEVVIVLYNEQSGALET	NLVSNKLGETEKSELIFKQEYSEK	TPEPHQPDTEKEK	
			: : : :		
610	620	630	640	650	
gi-19224134.	VLMGGQSESVEFTKD	TQMSGFSETVTIVEDTRPKLVFHFDNNEP	KVEEN	REK	
180	190	200	210		
gi-13621430.	PQKKRNGILPSTGEMVS	YVSVSALGIVLVATITLYSIYKKLKTSK			
	: : : : :	: : : :			
660	670	680	690		
gi-19224134.	PTKNITPILPATGDIENVLAFLGILILSVLSIFSLLKNQNNKV				

! Distributed over 1 thread.

! Start time: Wed Sep 15 18:45:36 2004

! Completion time: Wed Sep 15 18:45:38 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:00.3

! Total CPU time: 0:00:00.4

! Output File: gi-13621430.fasta

TYPE 3 pilus motifs / US605/27239
 protein F2 like fibronectin-binding protein
 Length: 696-733
 LPXTG
 pilin motif consensus PK (X₇) K
 E box consensus ETxAPxGY

SpyM3_0104/21909640	pilin motif	155 PKEKPIIYFK 398
	E box	YTFVETAAPDGY
SPs0106/28895018	pilin motif	269 PKEKPIIYFK 512
	E box	YTFVETAAPDGY
SpyM18_0132/19745307	pilin motif	269 PKEKPIIYFK 512
	E box	YTFVETAAPDGY
orf84	pilin motif	269 PKEKPIIYFK 512
	E box	YTFVETAAPDGY

TYPE 4 pilus motifs
 protein F2 like fibronectin-binding protein
 Length: 1161
 LPXTG
 pilin motif consensus PK (X₇₋₈) K
 E box consensus YxLxETxAPxGY

The protein is longer than the proteins belonging to TYPE 3 and has 4 possible pilin motifs and 2 E boxes

19224141	pilin motifs	215 PKGISQDIPK 571 PKGYQQVTEK 156 PKMSVVSKYGK 674 PKYDAKNQEYK
	E boxes	563 YDLYETKAPKGY 940 YTFVETAAPDGY

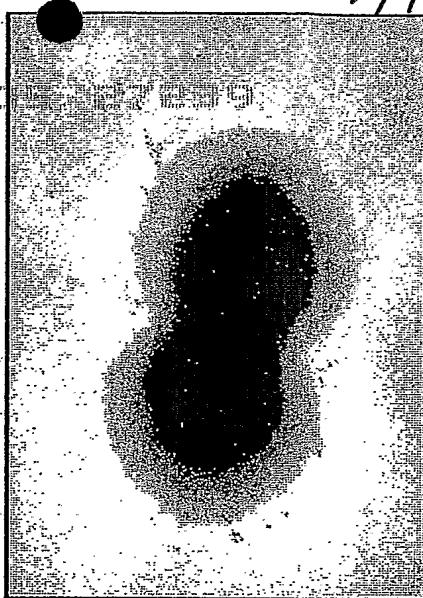
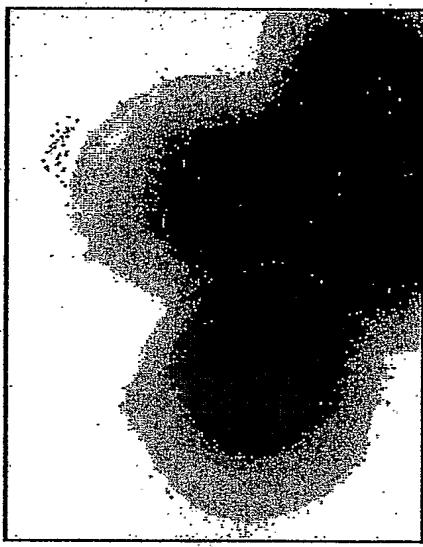
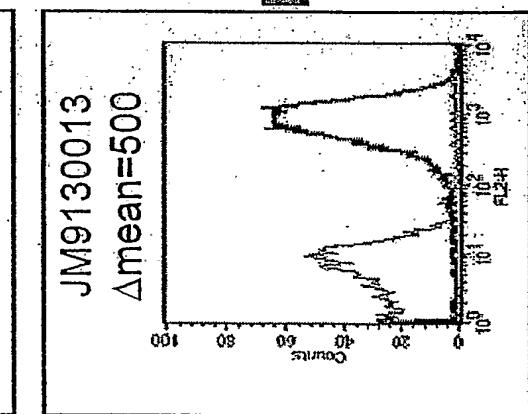
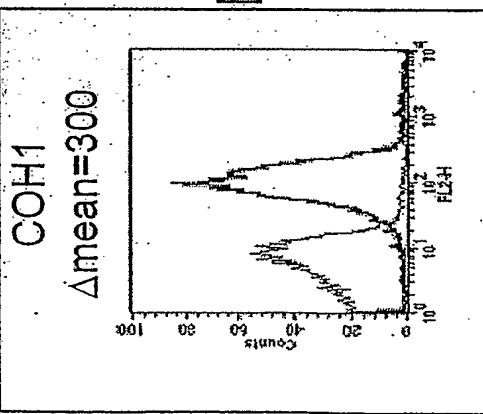
Formation of pili structures on GBS appears to be correlated to FACS values for surface expression of GBS80 protein

WO 2006/078318
PCT/US05/027239

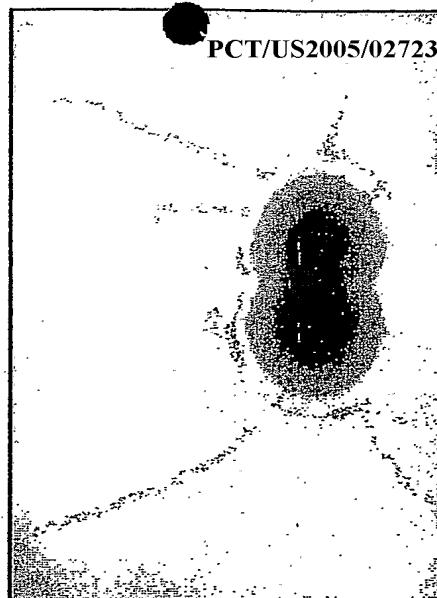
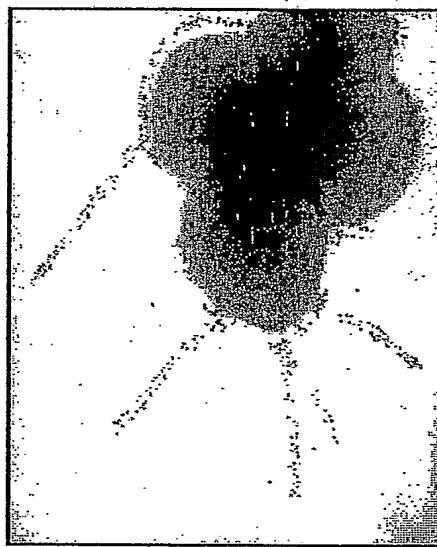
21/1/87

Immunogold Electron Microscopy

α -GBS80



α -GBS80



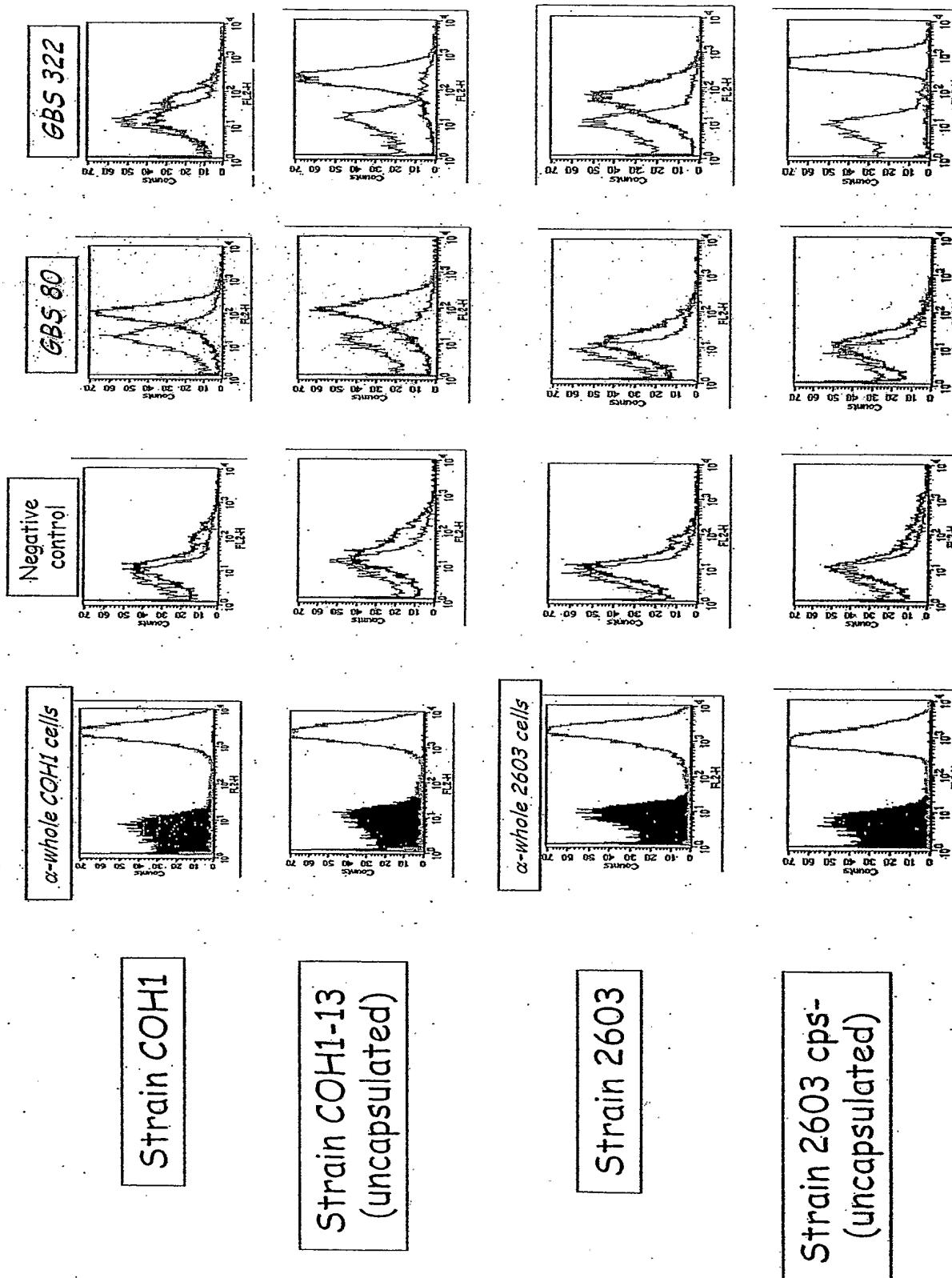
PCT/US2005/027239

CHIRON | VACCINES

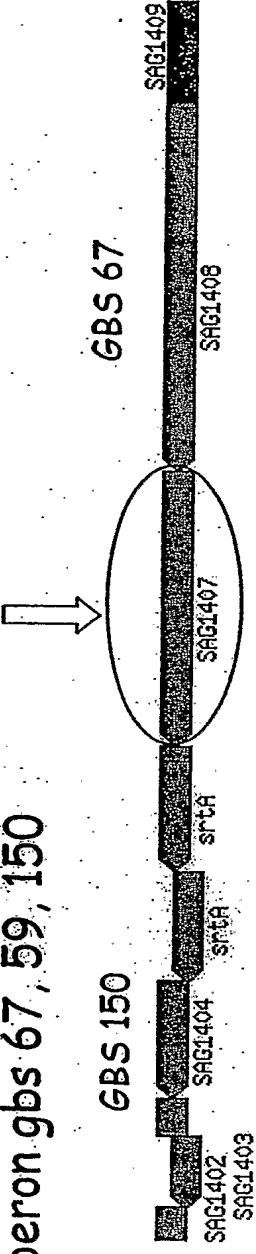
212/487

**Surface exposure is capsule-dependent
for GBS 322 but not for GBS 80**

FIGURE 62



**Adhesin island 2-
Operon gbs 67, 59, 150**

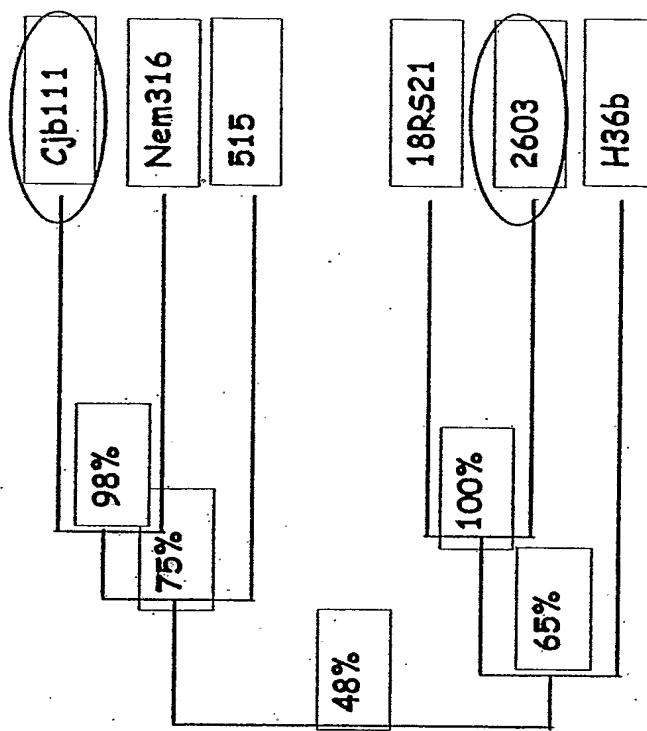


• By comparing GBS 59 amino acidic sequence of 2603 with that of other sequenced genomes, the following homologies are obtained:

2603	47%
nem316	62%
h36b	48%
515	48%
cjb111	100%
18rs21	not present (Sppb1)
coh1	not present (Sppb1)
A909	

• CGH: 1/20 GBS strains analyzed (18RS21)

Two-by-two amino acidic sequence comparison



There seem to be two clusters suggesting the presence of two major isoforms

FIGURE 63

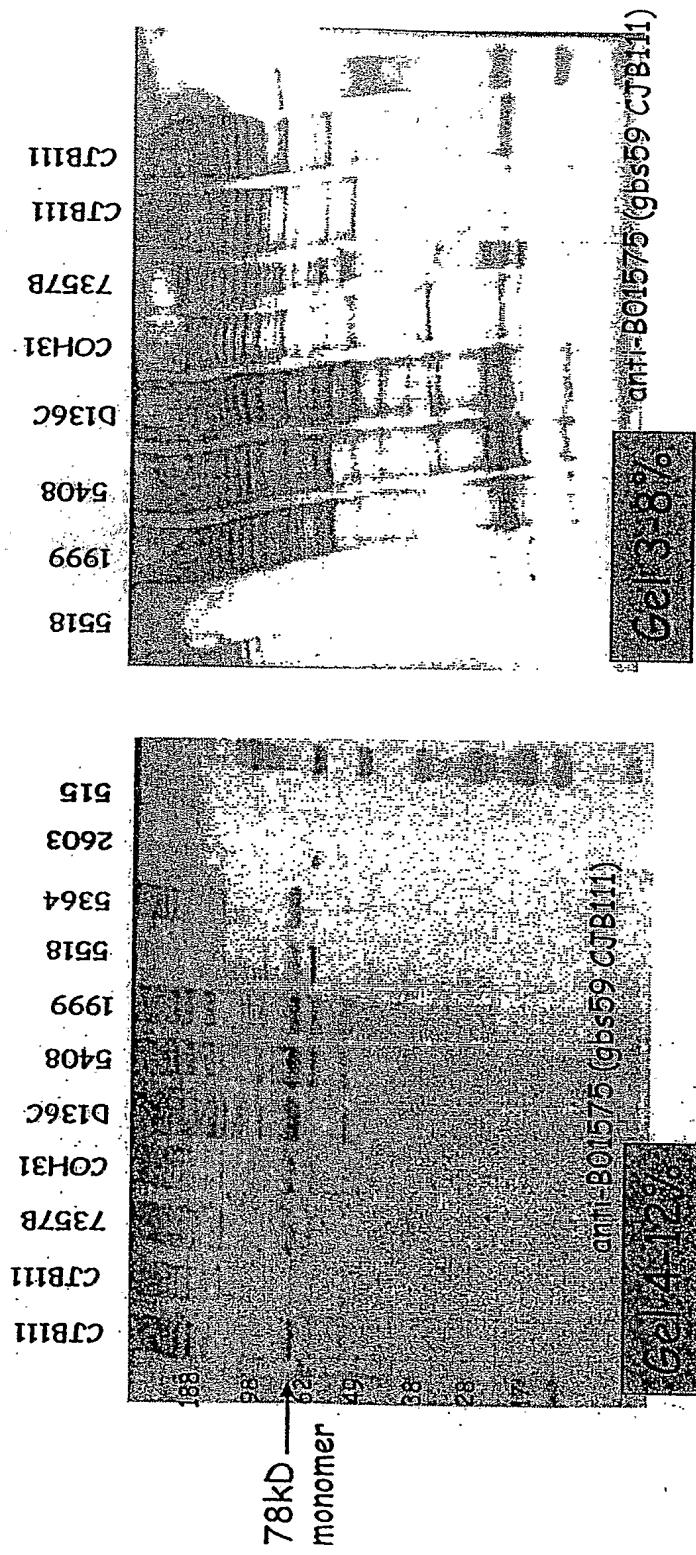
Western blotting with whole extracts
derived from GBS strains

WO 2006/078318

PCT/US05/27239

214/487

PCT/US2005/027239



GBS 59 is part of a high molecular weight polymer (pilus)
in GBS strains: CJB111, 7357b, coh31, d1363c, 5408, 1999,
5364, 5518, 515

FIGURE 64

Western blotting with purified proteins and whole extracts derived from GBS strains

WO 2006/078318

PCT/US05/27239

215/487

PCT/US2005/027239

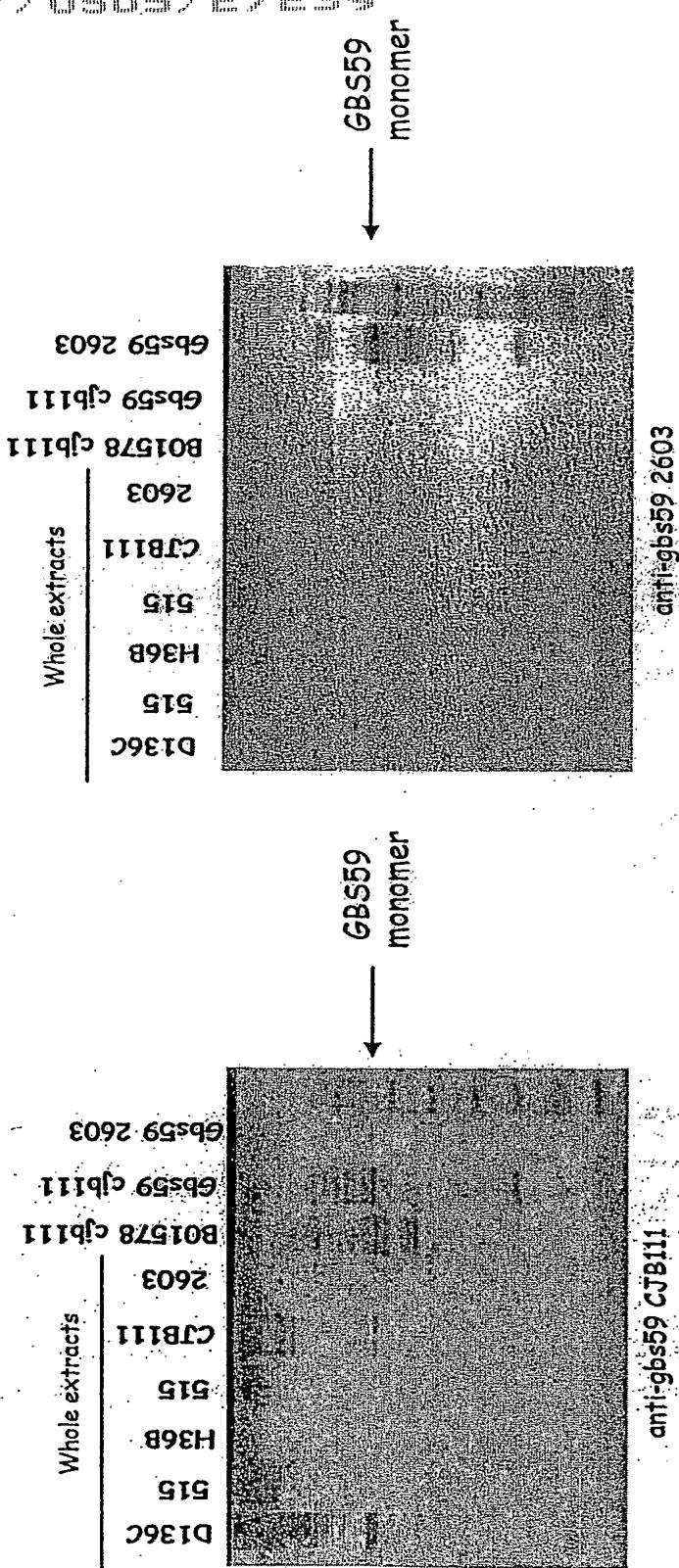


FIGURE 65

FACS analysis using mouse antiserum after immunization with BO1575 (gb559) from CJB111 genome

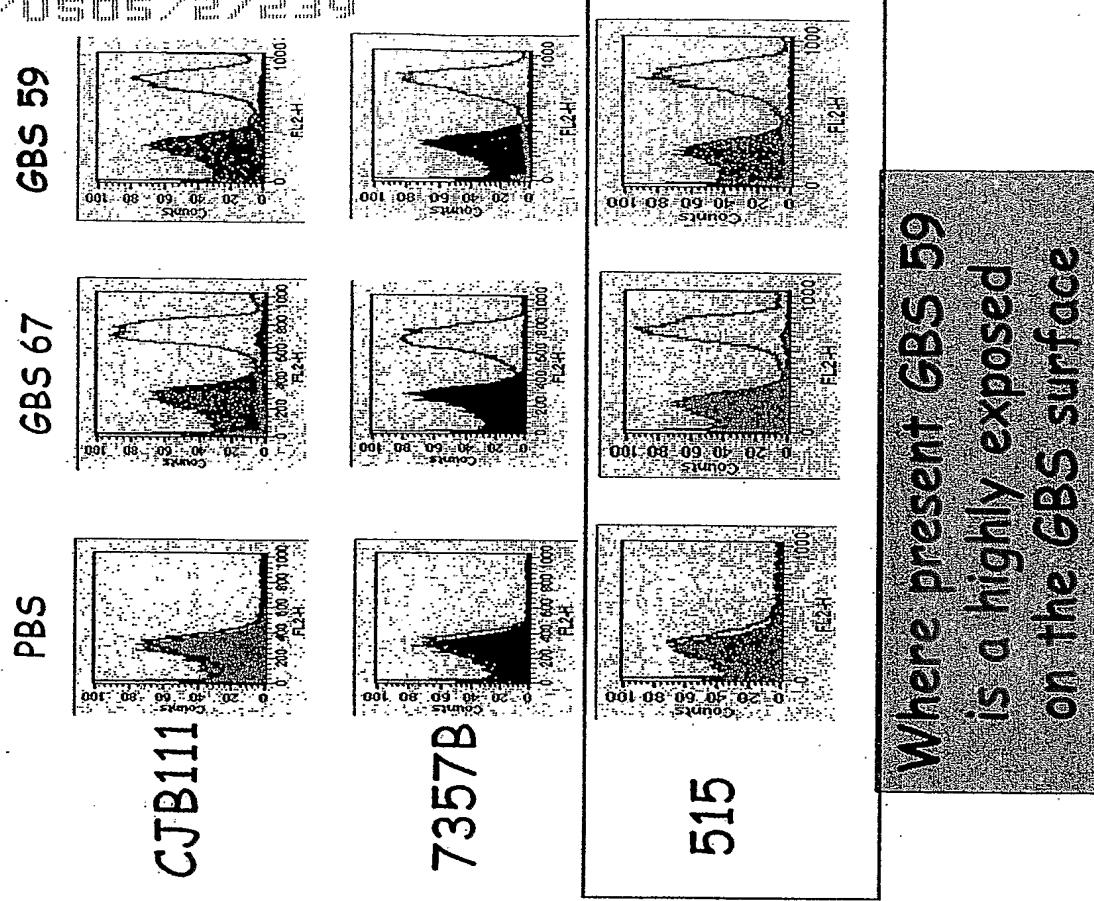
WO 2006/078318

216/487

PCT/US2005/027239

PCT/US05/27239

GBS strains	Type	FACS (Δ Mean)
DK1		565
DK8		559
Davis	Ia	577
5115		583
2986		443
5551		524
7357b-		596
5518		190
D136C		504
C0H31	III	505
DK21	II	249
CJB111		493
5364	V	593
2110		590
1999		594
2210		536
5408		567
1169		227



Where present GBS 59
is a highly exposed
on the GBS surface

FIGURE 66

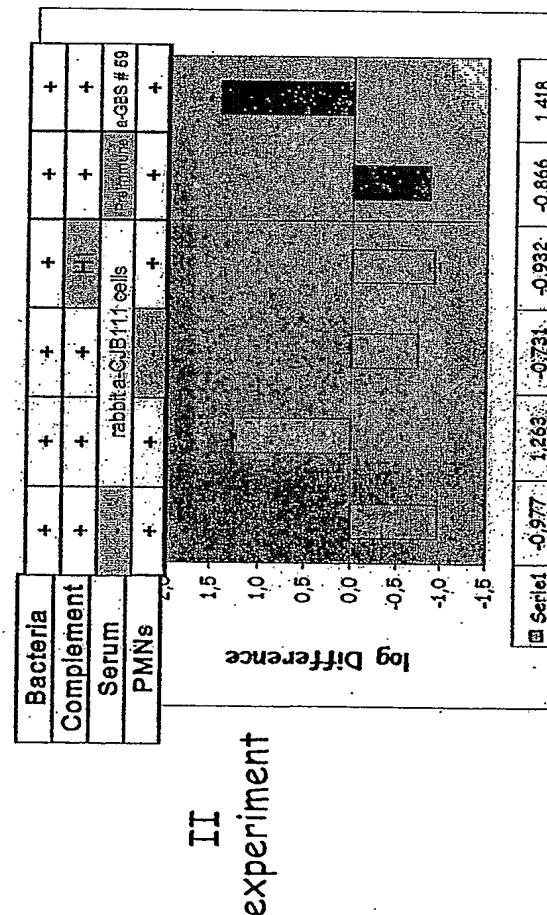
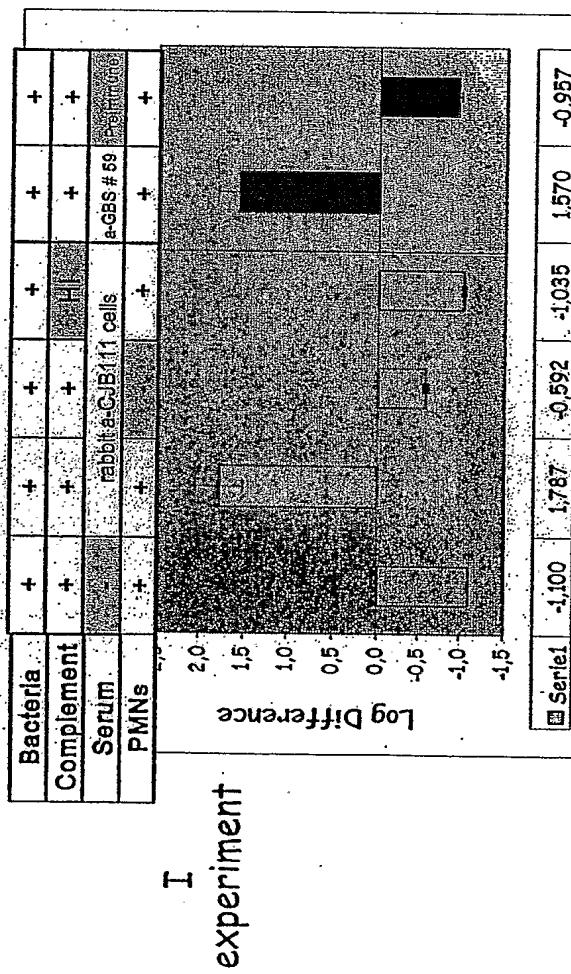
Opsonophagocytosis assays: B01575 (gbss59-cjb111)

WO 2006/078318

PCT/US05/27239

217/487

PCT/US2005/027239



• CJB111 GBS strain type Ia
 • Baby rabbit complement
 • Human PMNs
 • Positive control: anti-type V cells
 (rabbit serum anti fixed type V cells)

FIGURE 67

Association GBS 80-104

WB a-80, a-104 JM9130013 Total Extract

WO 2006/078318

PCT/US05/27239

218/487

PCT/US2005/027239

FACS	(ΔMEAN)
GBS 80	597
GBS 104	446

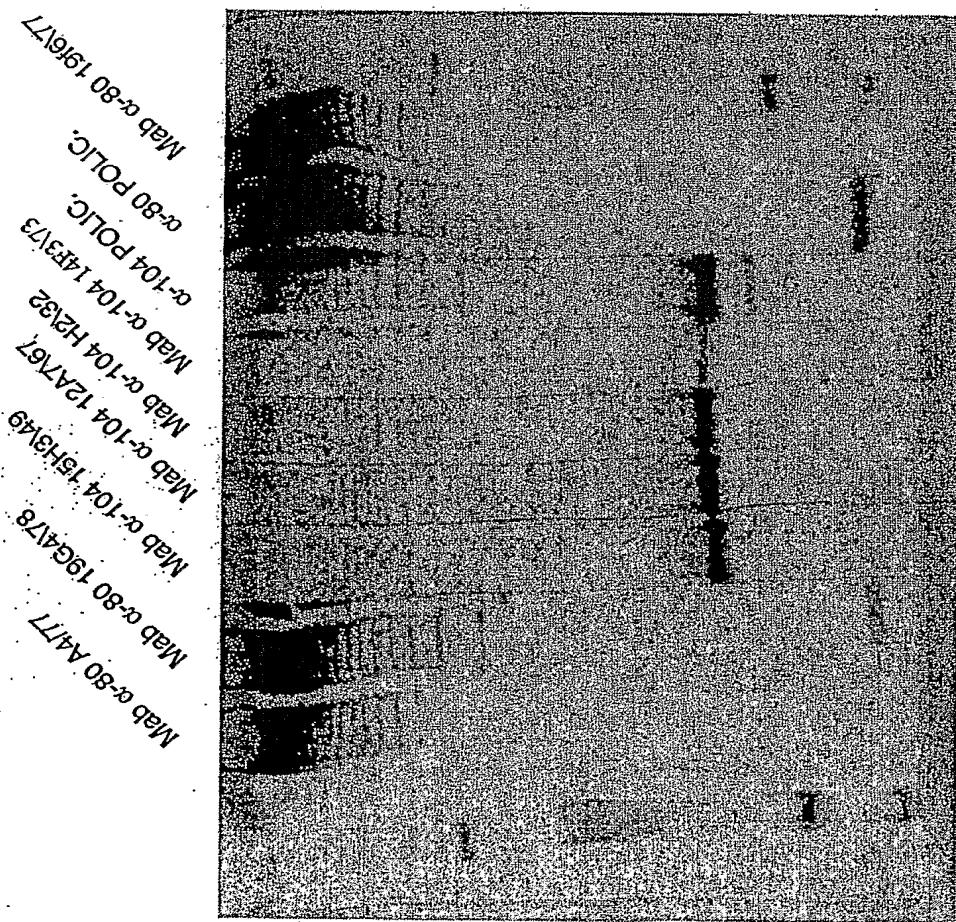
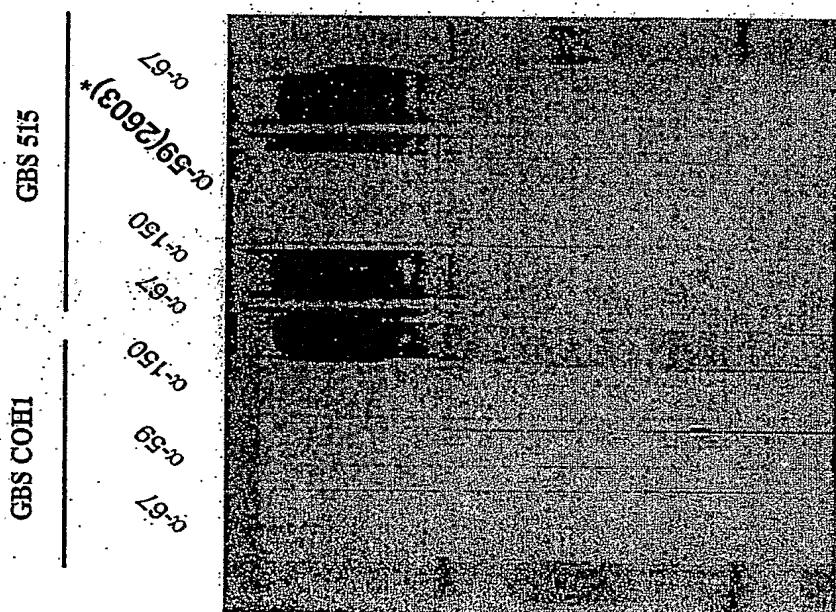


FIGURE 68

**WB GBS 515 Total Extract
 α -67; α -150; α -59 (2603)**



Controls:
GBS COH1 total extracts

- anti-gbs59 mouse serum after immunization with SAG1407 (GBS 59) from 2603 genome

GBS 67 and GBS 150 are parts of a high molecular weight polymer (pilus) in 515 GBS strain

Western Blotting ko GBS67
from 515 genome (clone 1.45)

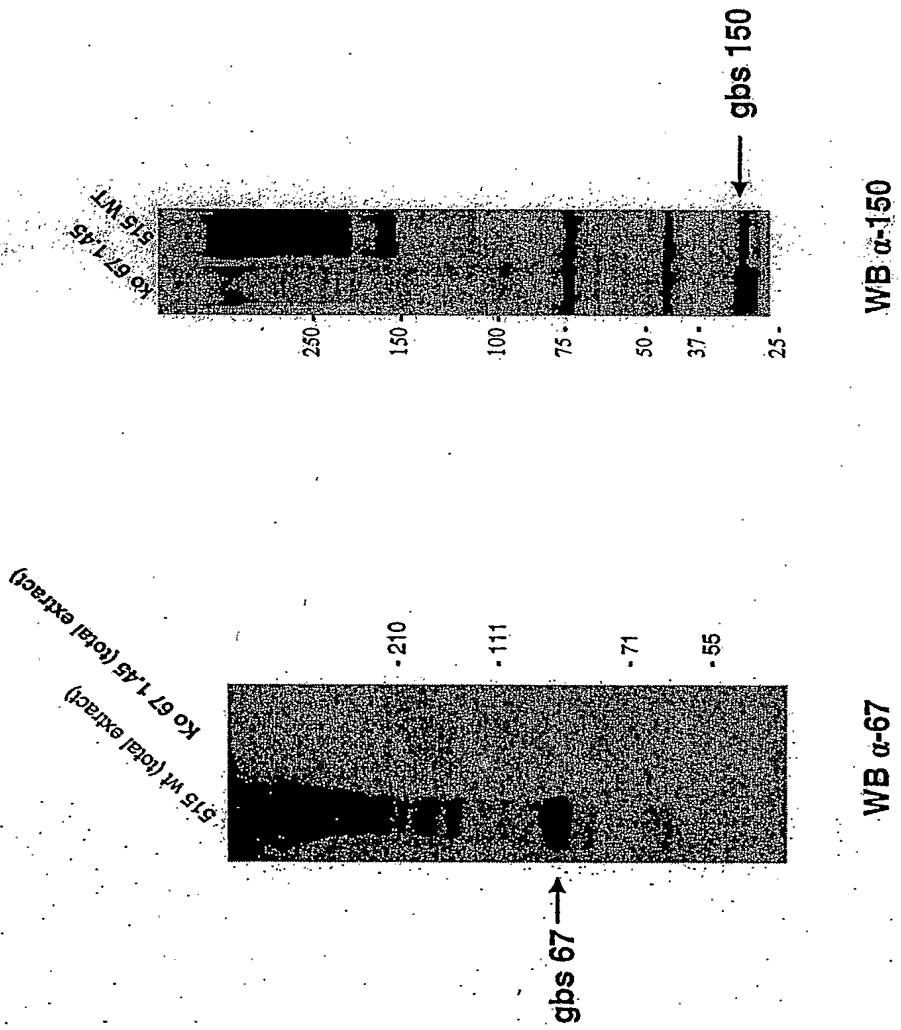


FIGURE 70

WO 2006/078318

220/487

PCT/US2005/027239

PCT/US05/27239

FACS GBS 515 Δ67

WO 2006/078318

PCT/US05/27239

221/487

PCT/US2005/027239

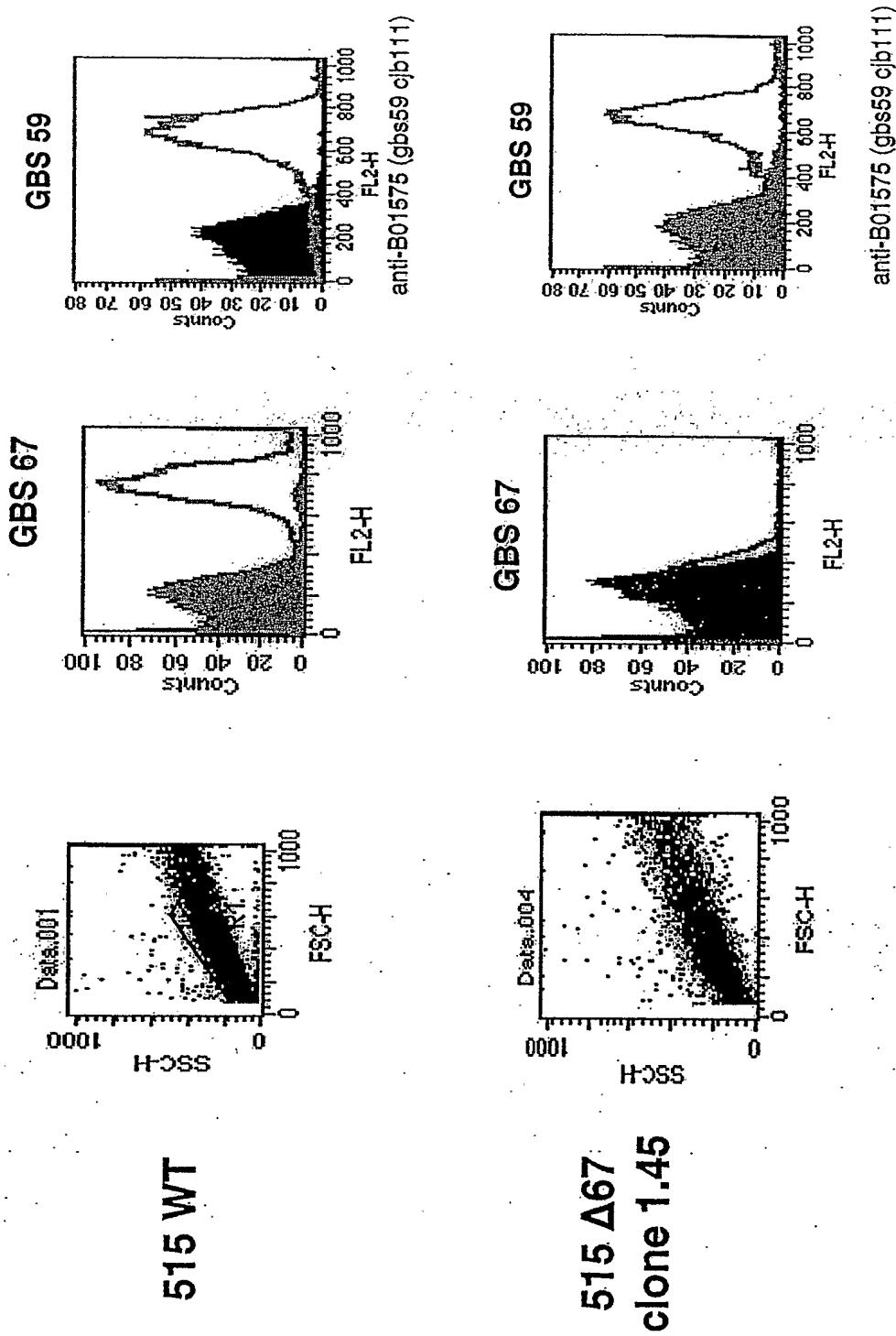


FIGURE 71

Complementation of GBS 515 KO 67 with pAM401-gbs80

WO 2006/078318

PCT/US05/27239

222/487

PCT/US2005/027239

GBS 80 forms a high molecular weight complex
(pilus) in absence of GBS 67

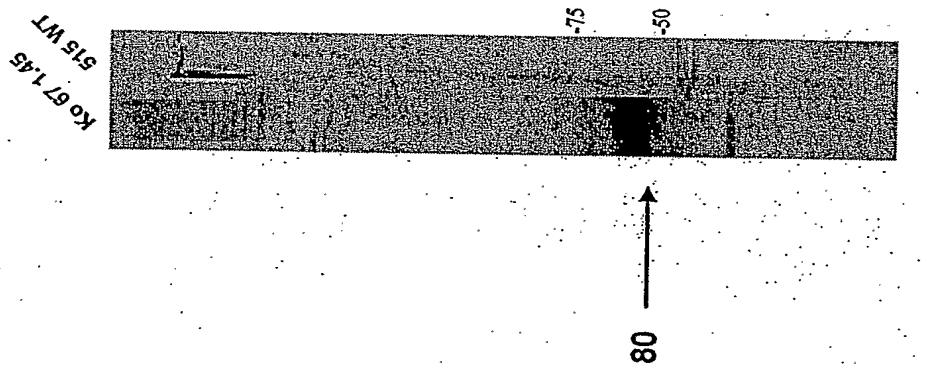


FIGURE 72

spyM6_0159 type I pilus present in M6

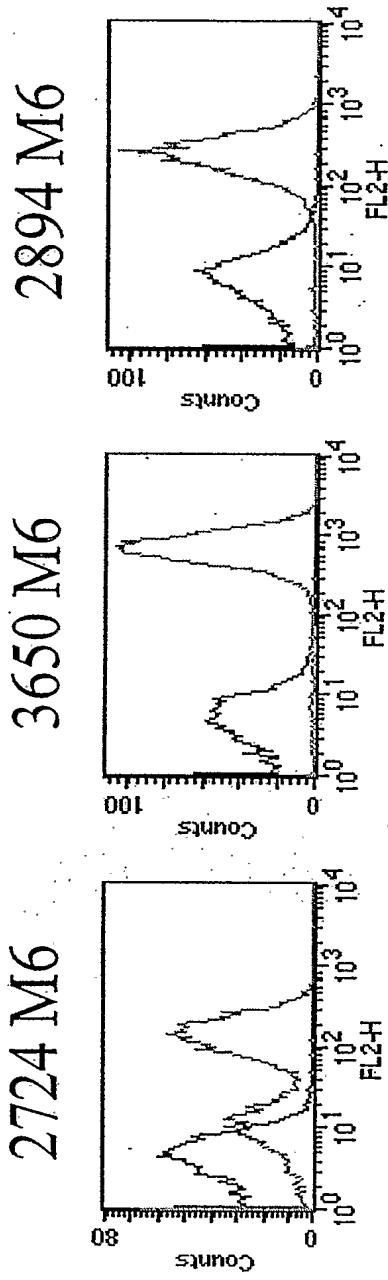


Figure 73

spyM6_0160 type 1 pilus present in M6

WO 2006/078318

PCT/US05/27239 224/487

PCT/US2005/027239

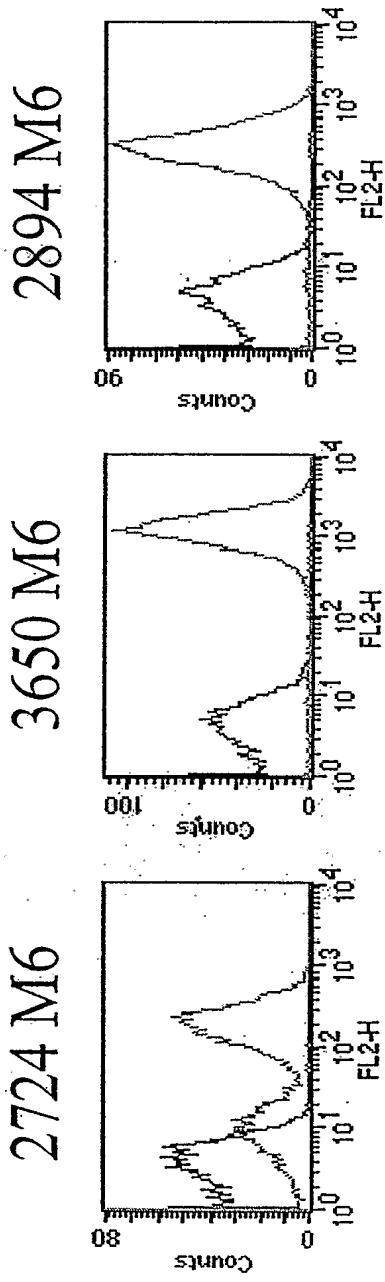


Figure 74

225/487

Gas15 type 2 pilus present in M1

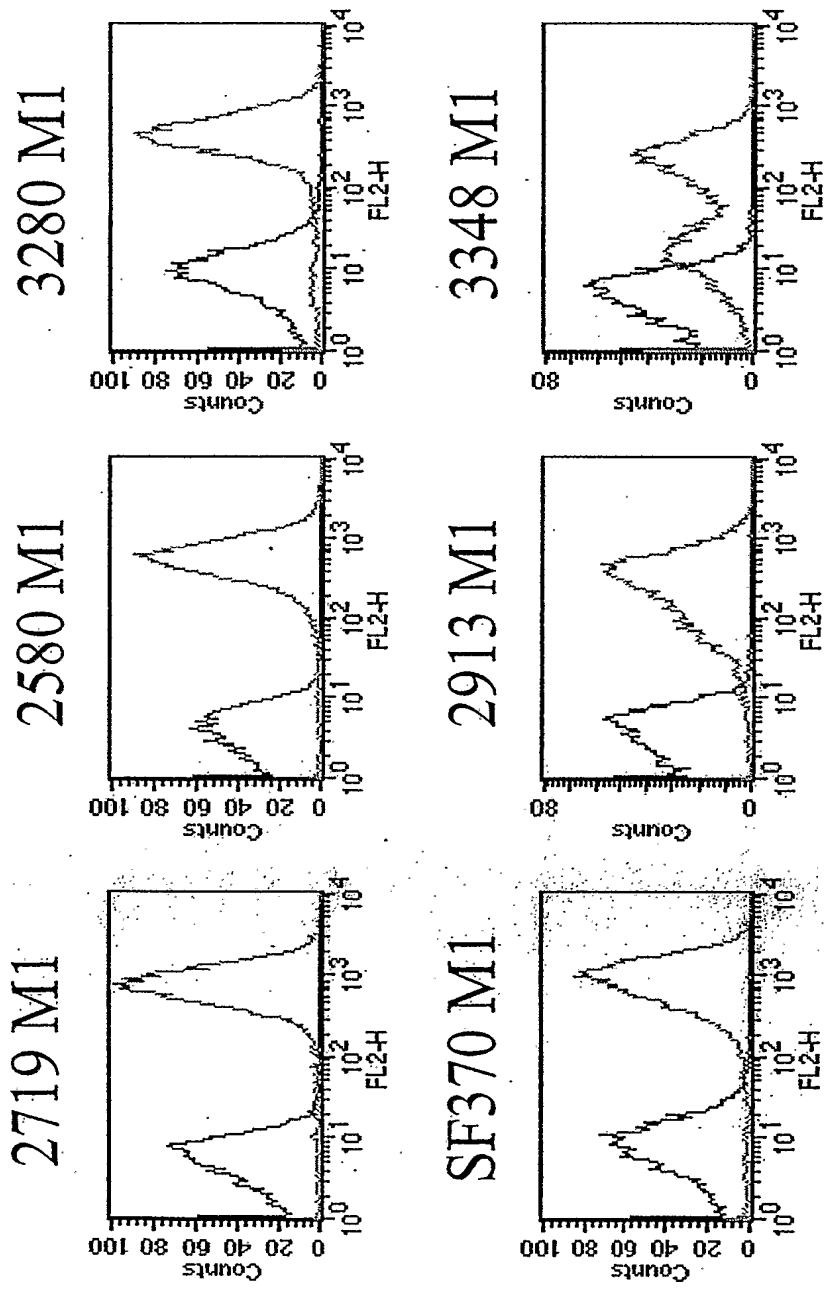


Figure 75

226/487

Gas16 type 2 pilus present in M1

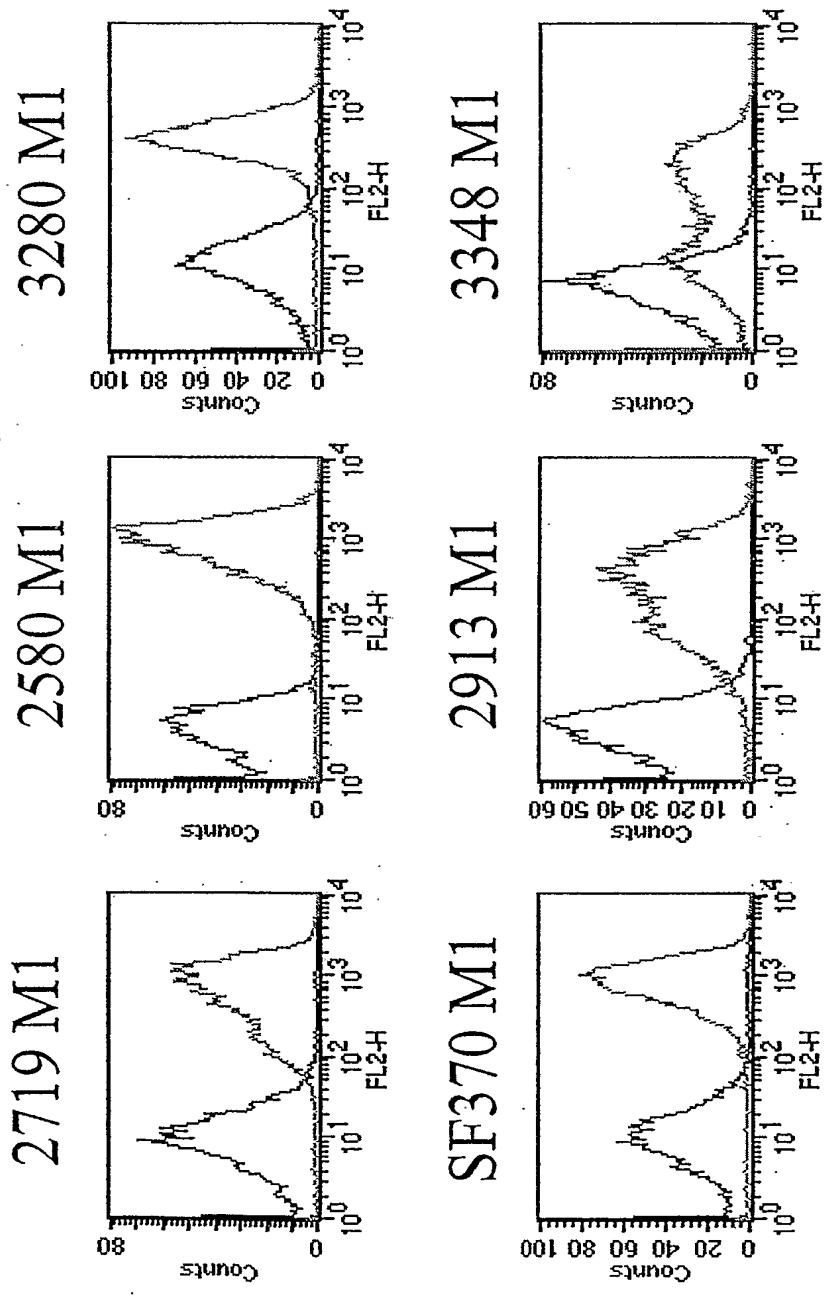


Figure 76

227/487

Gas18 serum 1 type 2 pilus present in M1

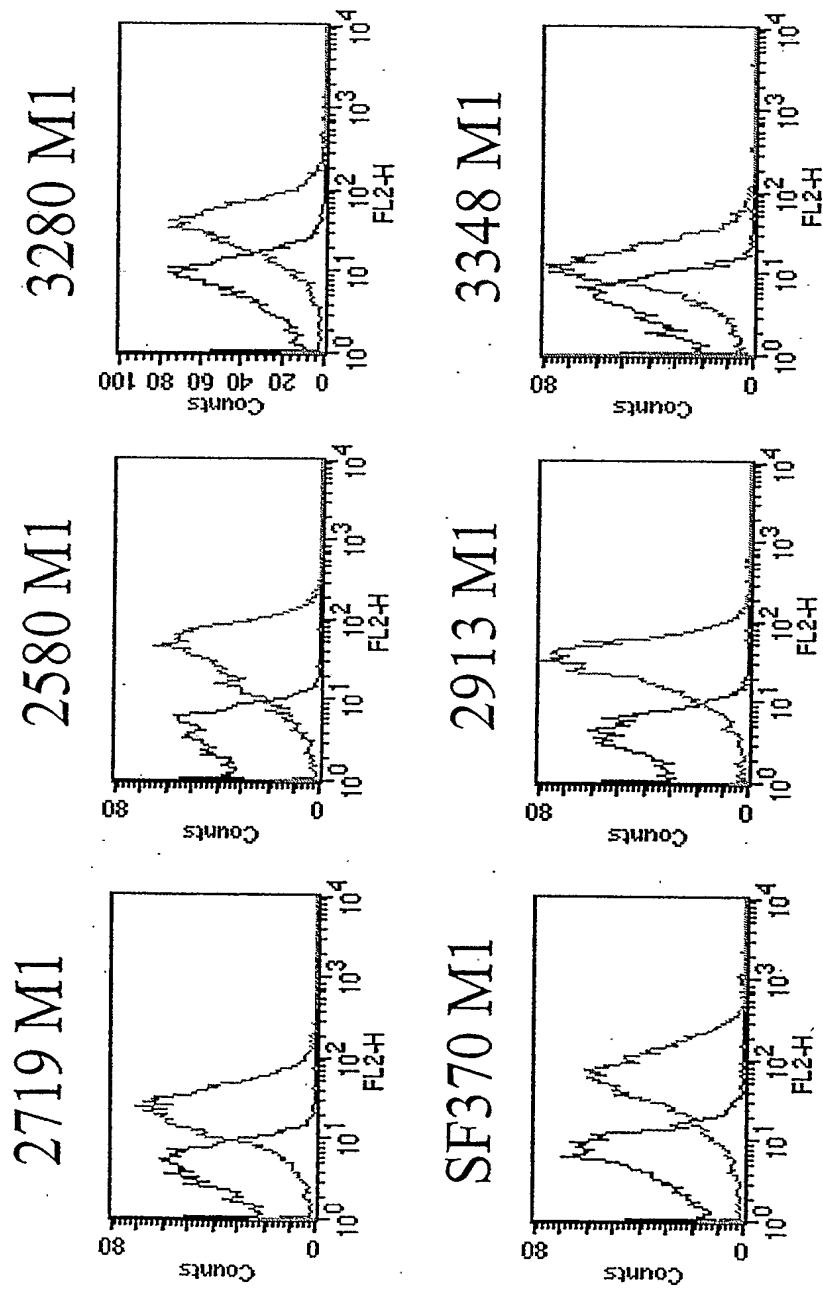


Figure 77

228/487

Gas18 serum 2 type 2 pilus present in M1

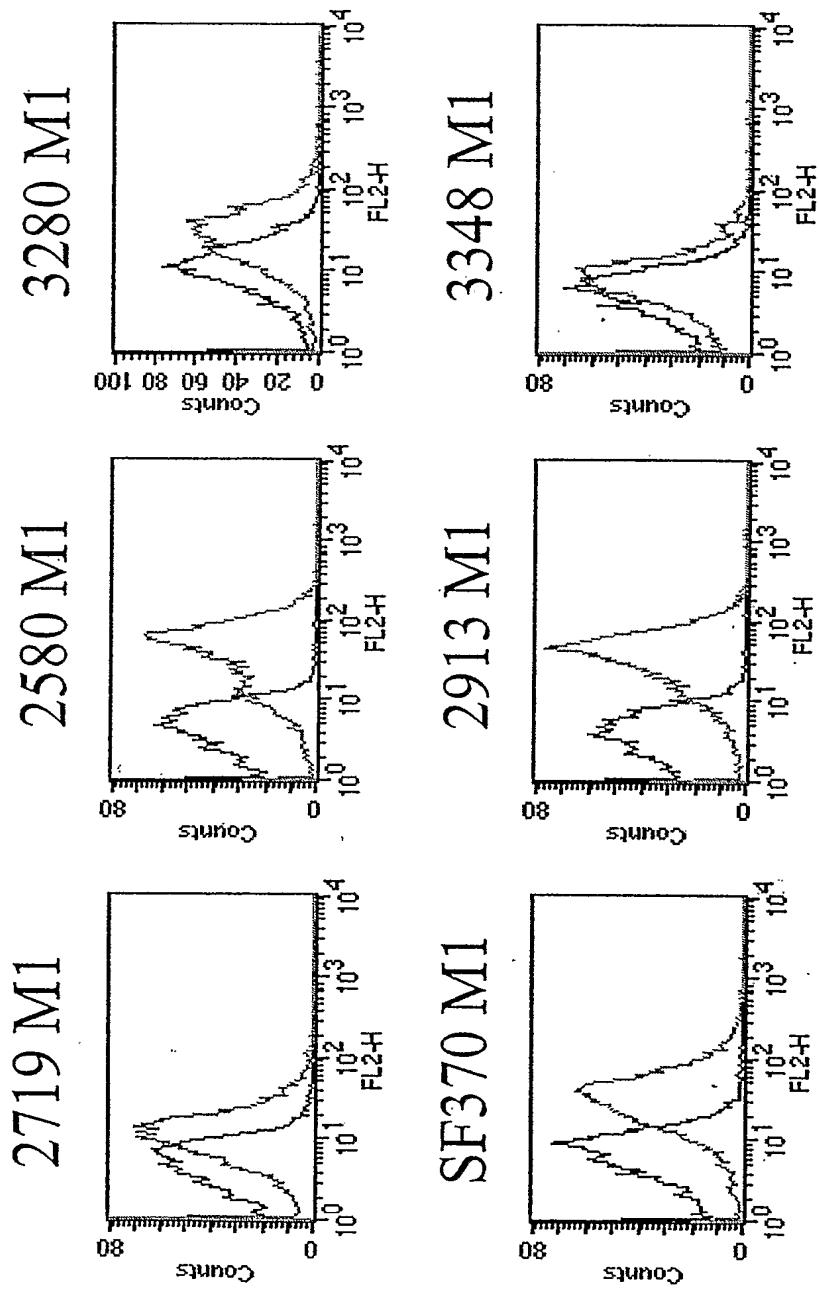


Figure 78

Gas16p2 type 2 pilus present in M1

WO 2006/078318

PCT/US2005/27239

PCT/US2005/027239

229/487

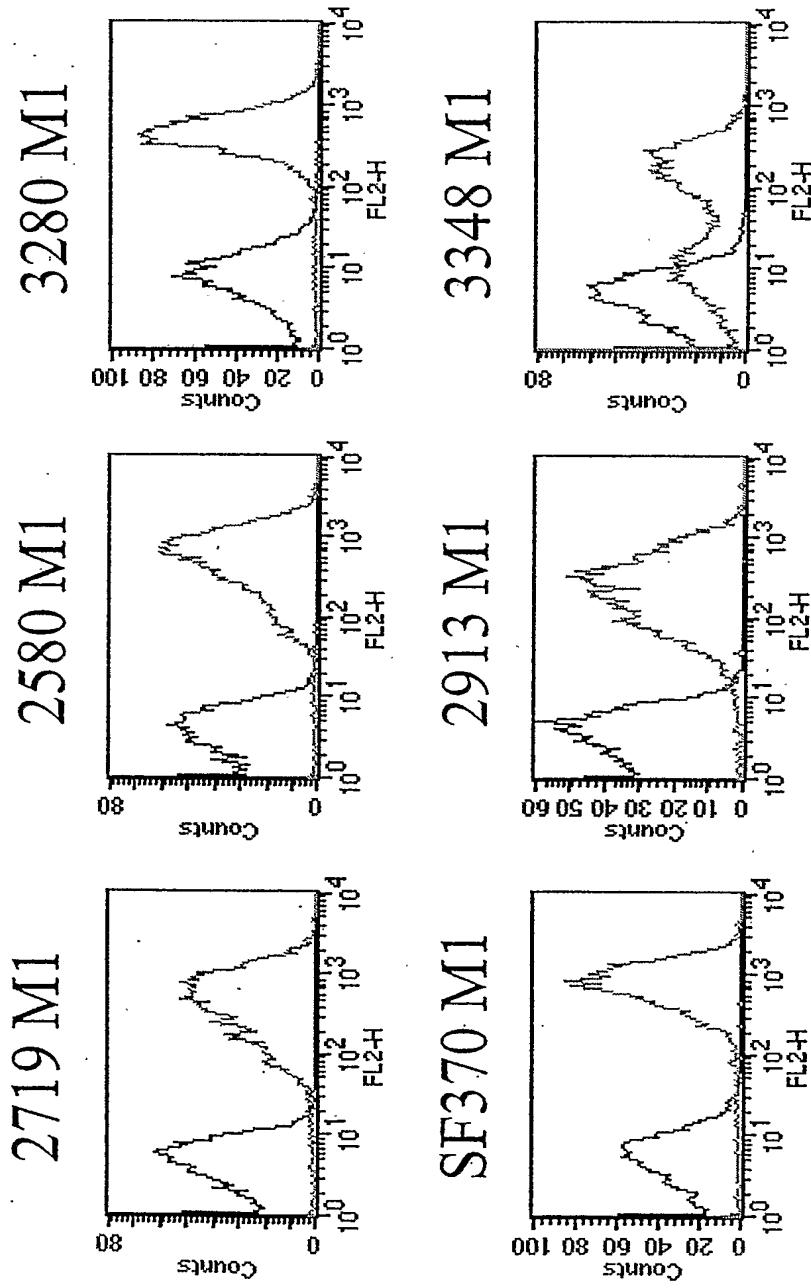


Figure 79

230/487

spyM3_0098 type 3 pilus present in M3

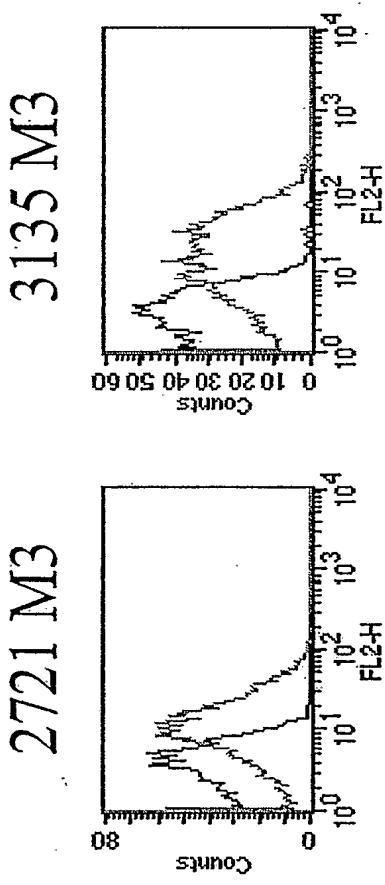
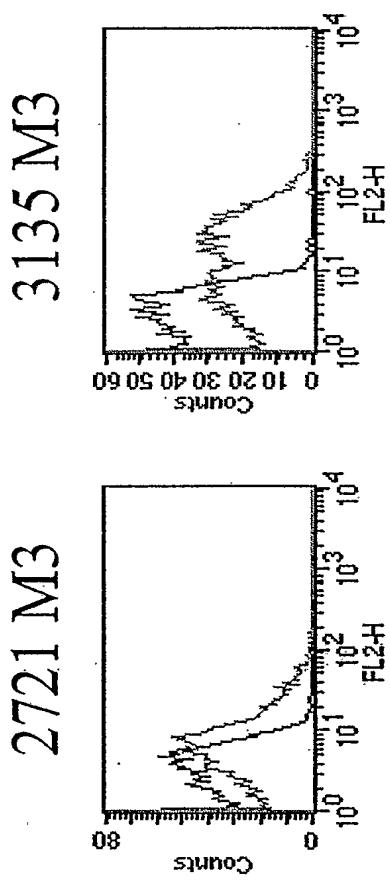


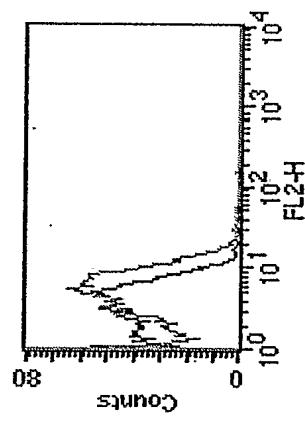
Figure 80

231/487

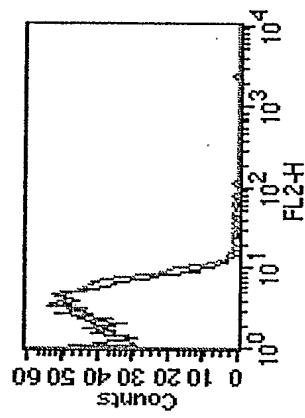
spyM3_0100 type 3 pilus present in M3**Figure 81**

spyM3_0102 type 3 pilus present in M3

2721 M3



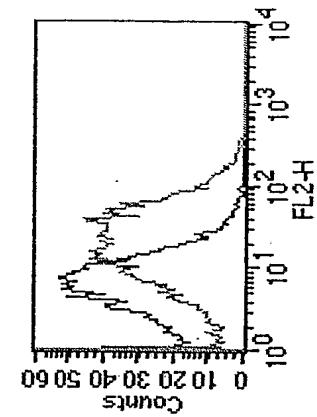
3135 M3



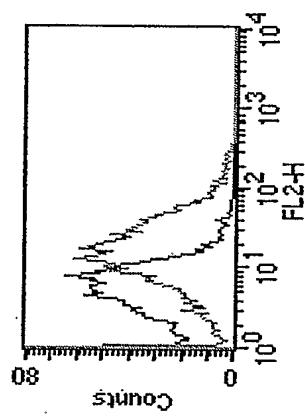
232/487

PCT/US2005/027239

2894 M6



3650 M6



2724 M6

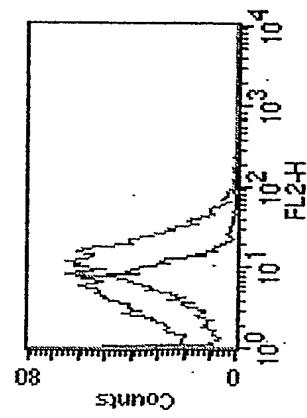
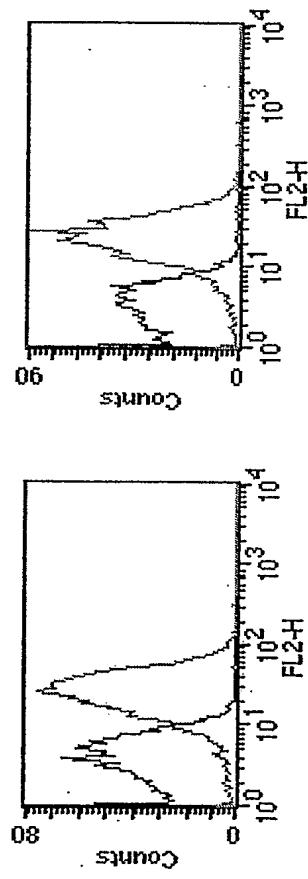
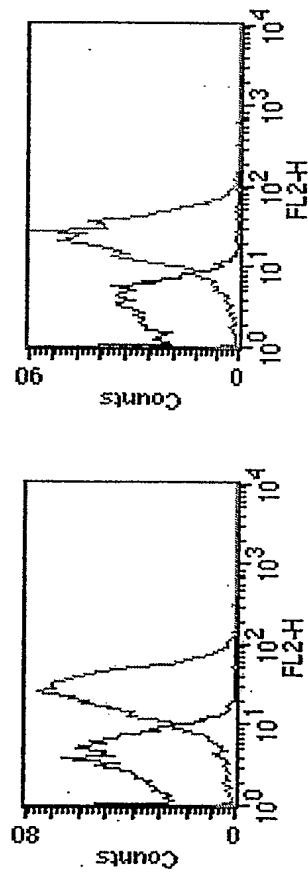
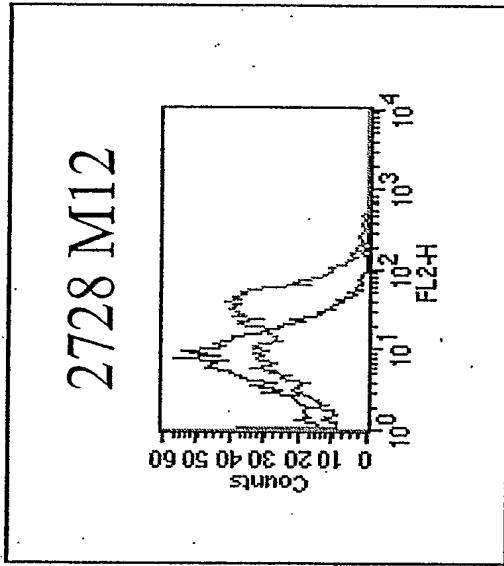


Figure 82

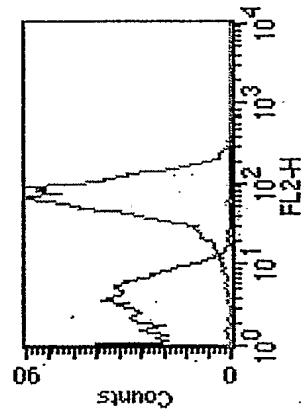
233/487
PCT/US2005/27239**spyM3_0104 type 3 pilus present in M3****2721 M3****3135 M3****2728 M12****Figure 83**

PCT/US05/27239

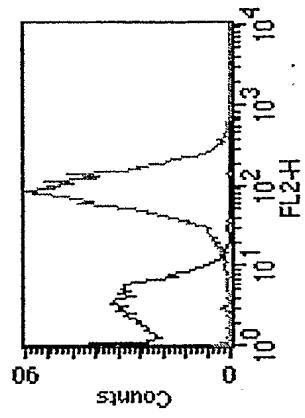
234/487

spyM3_0106 type 3 pilus present in M3

2721 M3



3135 M3



2728 M12

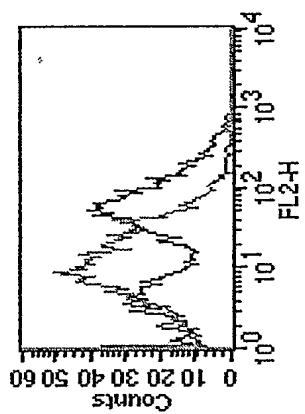
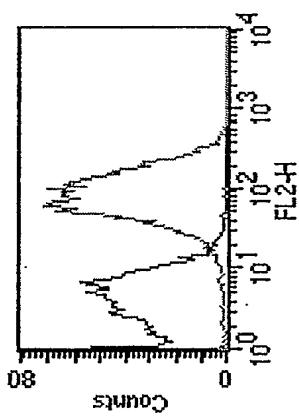


Figure 84

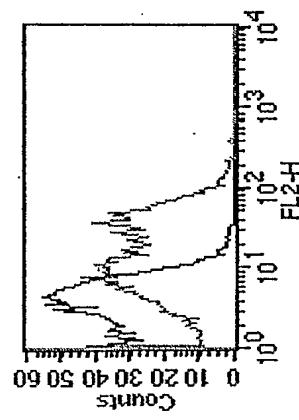
PCT/US05/27239 235/487

19224134 type 4 pilus present in M12

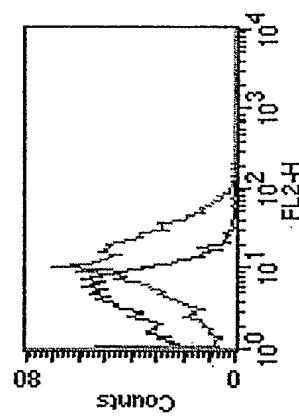
2728 M12



2724 M6



3650 M6



2894 M6

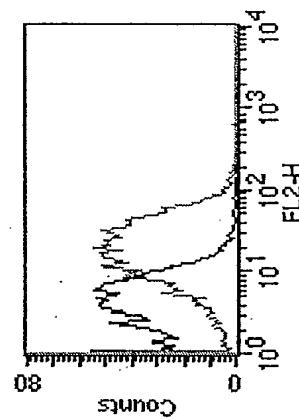


Figure 85

19224135 type 4 pilus present in M12

2728 M12

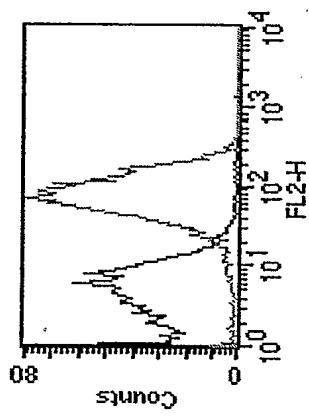


Figure 86

237/487

19224137 type 4 pilus present in M12

2728 M12

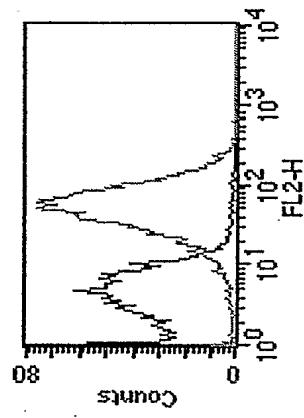


Figure 87

19224141 type 4 pilus present in M12

2728 M12

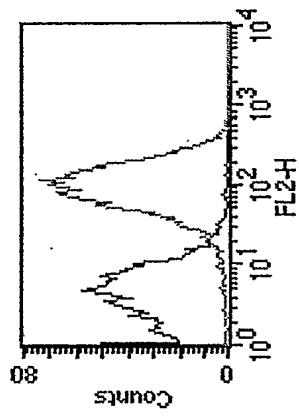


Figure 88

239/487

Figure 89

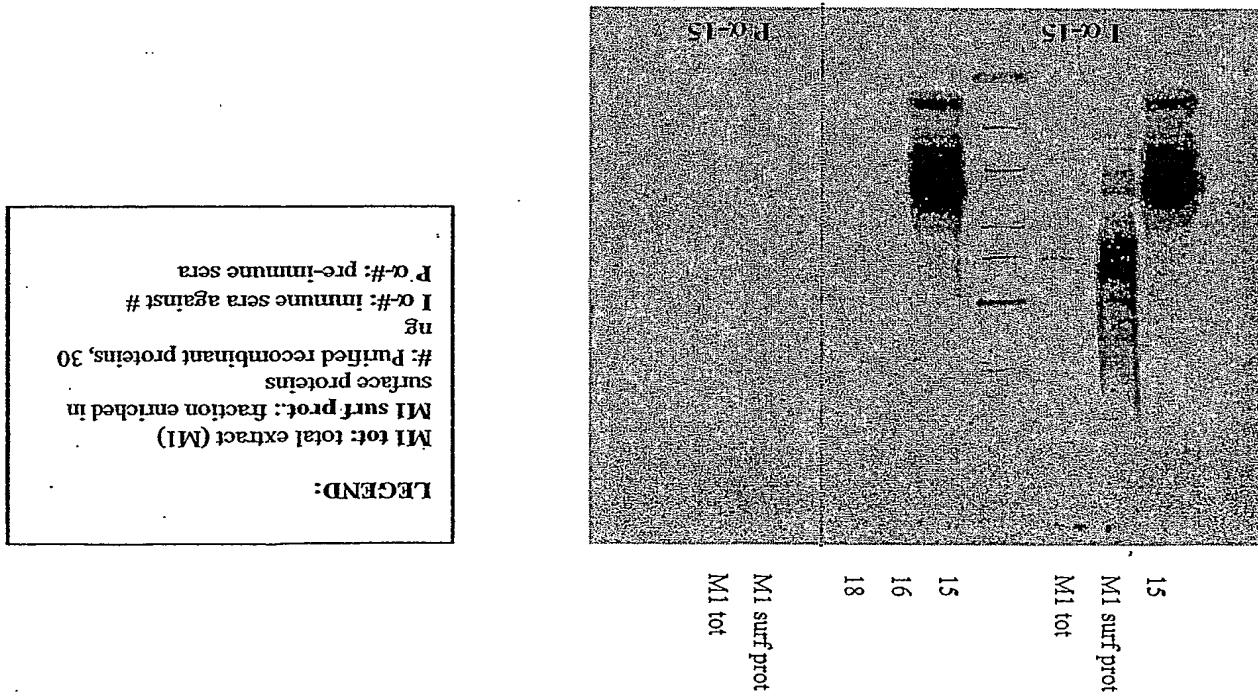
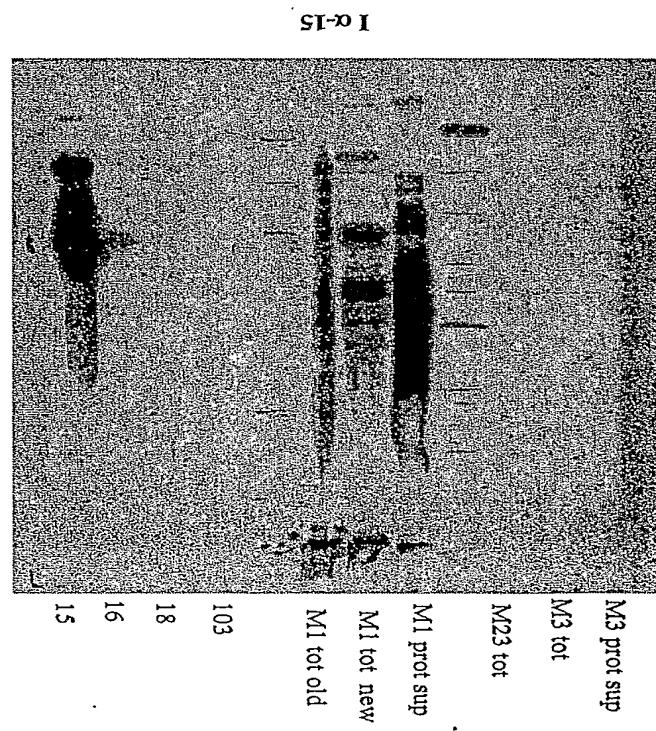
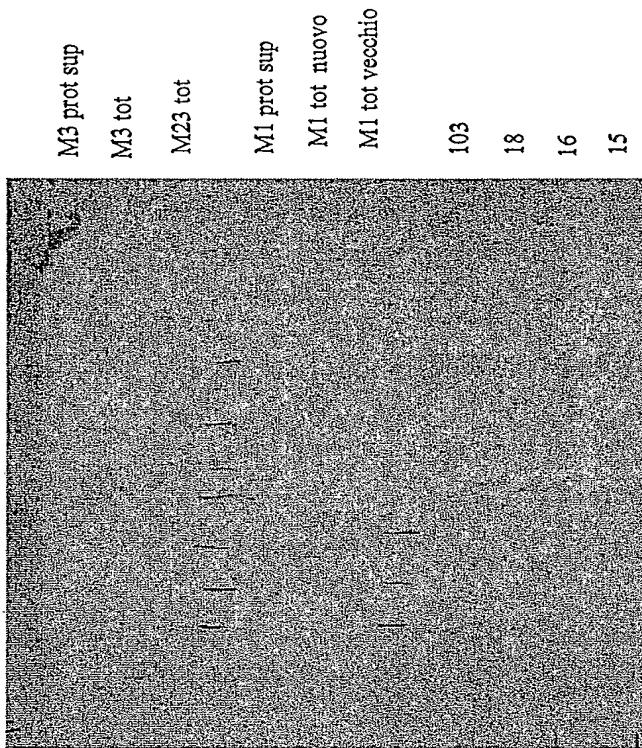


Figure 90

LEGENDA:

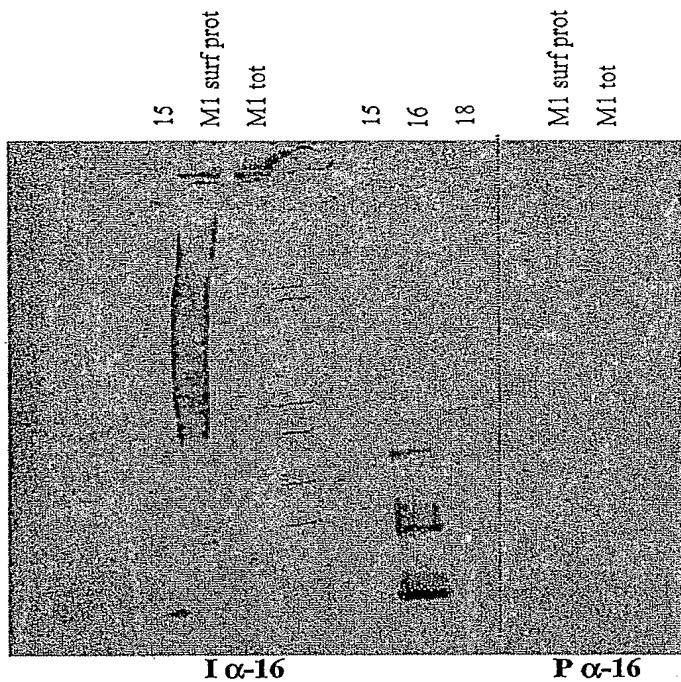
M1 tot total extract (M1)
 M1 prot Sup: fraction enriched in surface proteins
 #: Purified recombinant proteins, 30 ng
 Ig#: immune sera against #
 P #: pre-immune sera



PCT/US05/27239
241/487 $P \alpha-15$ **LEGENDA:**

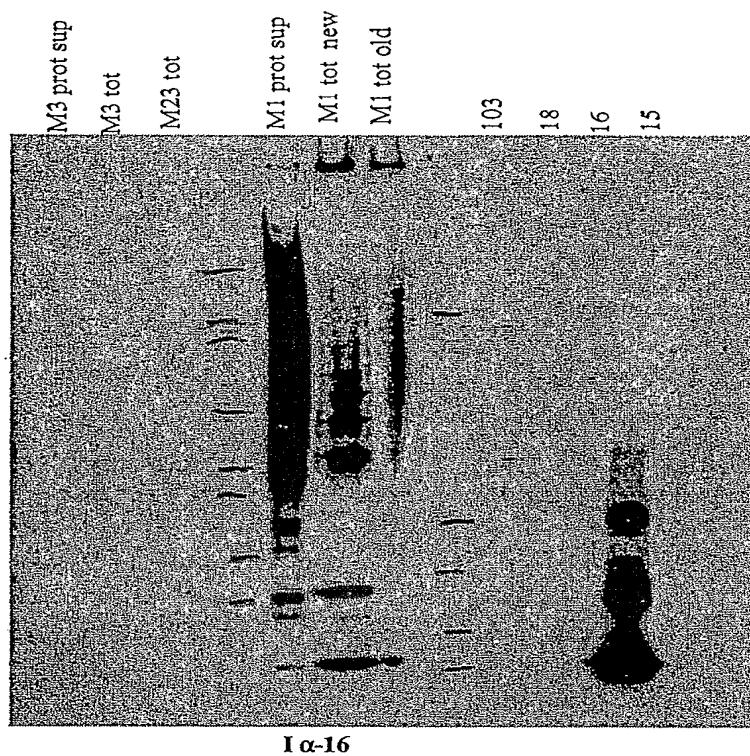
- M1 tot:** total extract (M1)
- M1 prot. Sup.:** fraction enriched in surface proteins
- #: Purified recombinant proteins, 30 ng
- I α -#:** immune sera against #
- P α -#:** pre-immune sera

Figure 91

**LEGEND:**

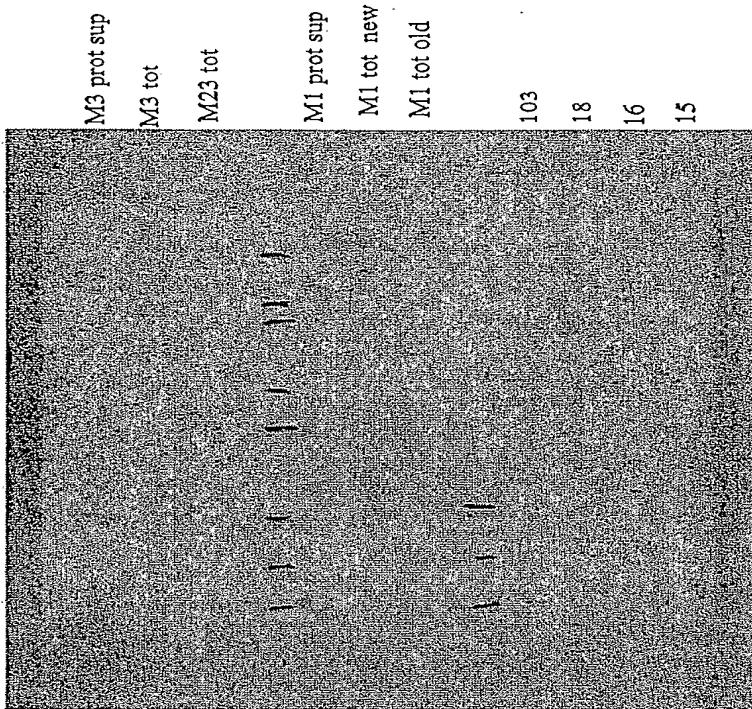
M1 tot: total extract (M1)
M1 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α -#: immune sera against #
P α -#: pre-immune sera

Figure 92

**LEGENDA:**

M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
 $I \alpha-\#$: immune sera against #
 $P \alpha-\#$: pre-immune sera

Figure 93

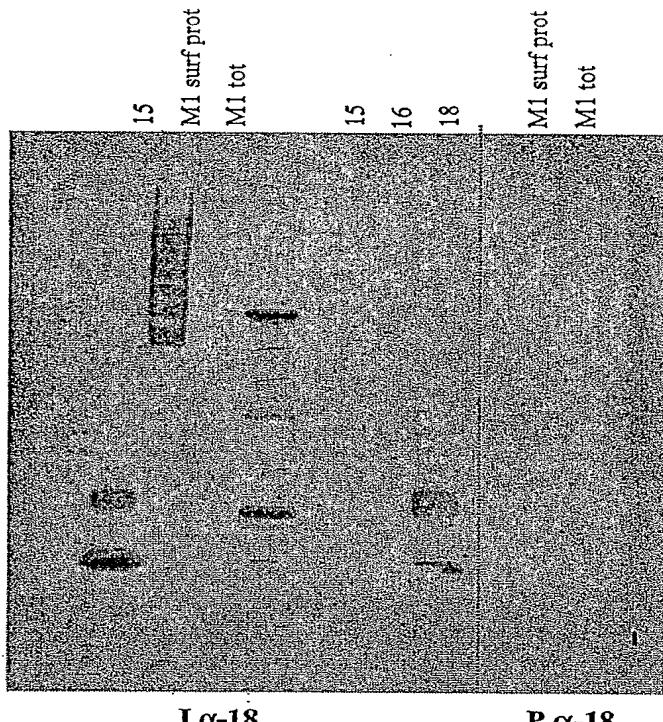


P α-16

LEGENDA:

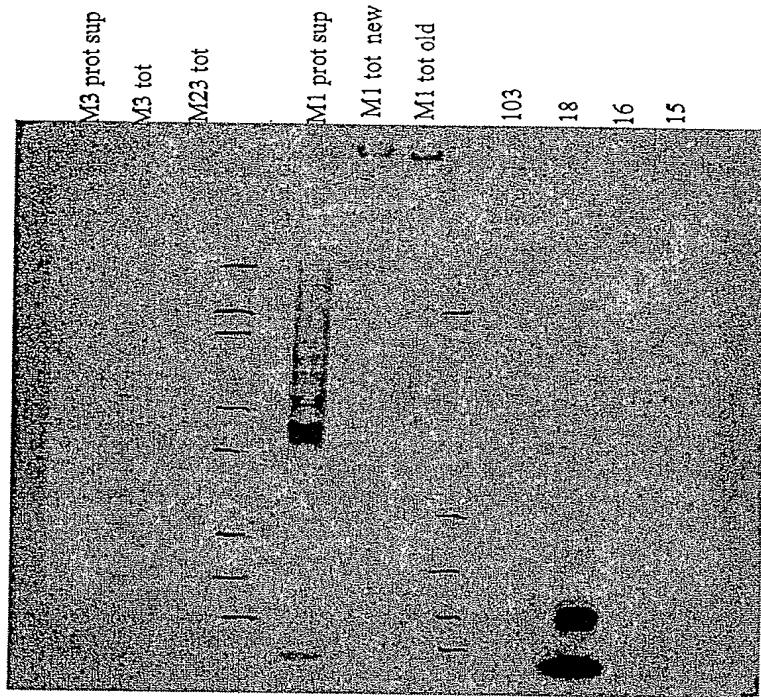
- M1 tot:** total extract (M1)
- M1 prot. Sup.:** fraction enriched in surface proteins
- #: Purified recombinant proteins, 30 ng
- I α-#:** immune sera against #
- P α-#:** pre-immune sera

Figure 94

**LEGEND:**

M1 tot: total extract (M1)
M1 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
 $I \alpha\text{-}#$: immune sera against #
 $P \alpha\text{-}#$: pre-immune sera

Figure 95

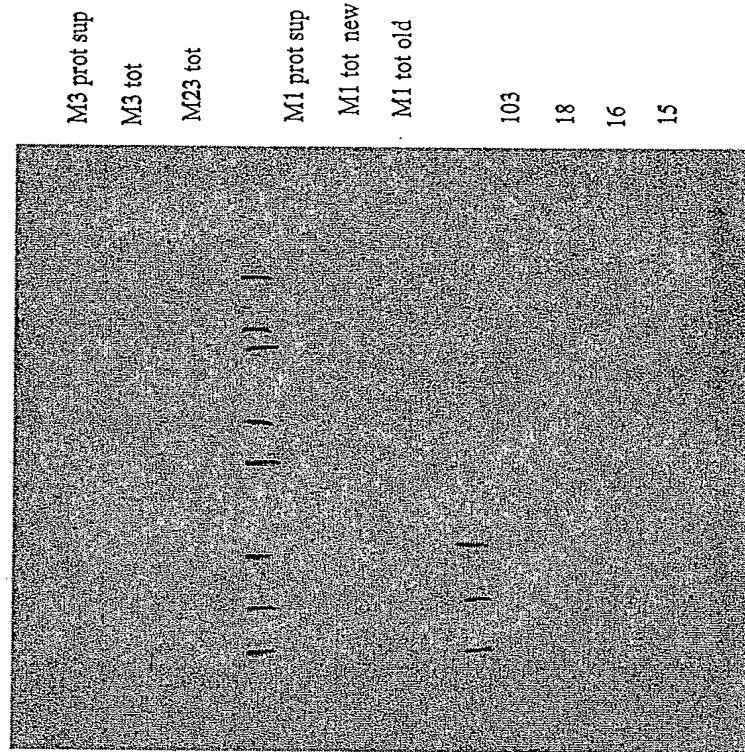


$I \alpha\text{-}18$

LEGENDA:

- M1 tot: total extract (M1)
- M1 prot. Sup.: fraction enriched in surface proteins
- #: Purified recombinant proteins, 30 ng
- I $\alpha\text{-}\#$: immune sera against #
- P $\alpha\text{-}\#$: pre-immune sera

Figure 96

PCT/US05/27239
247/487**LEGENDA:**

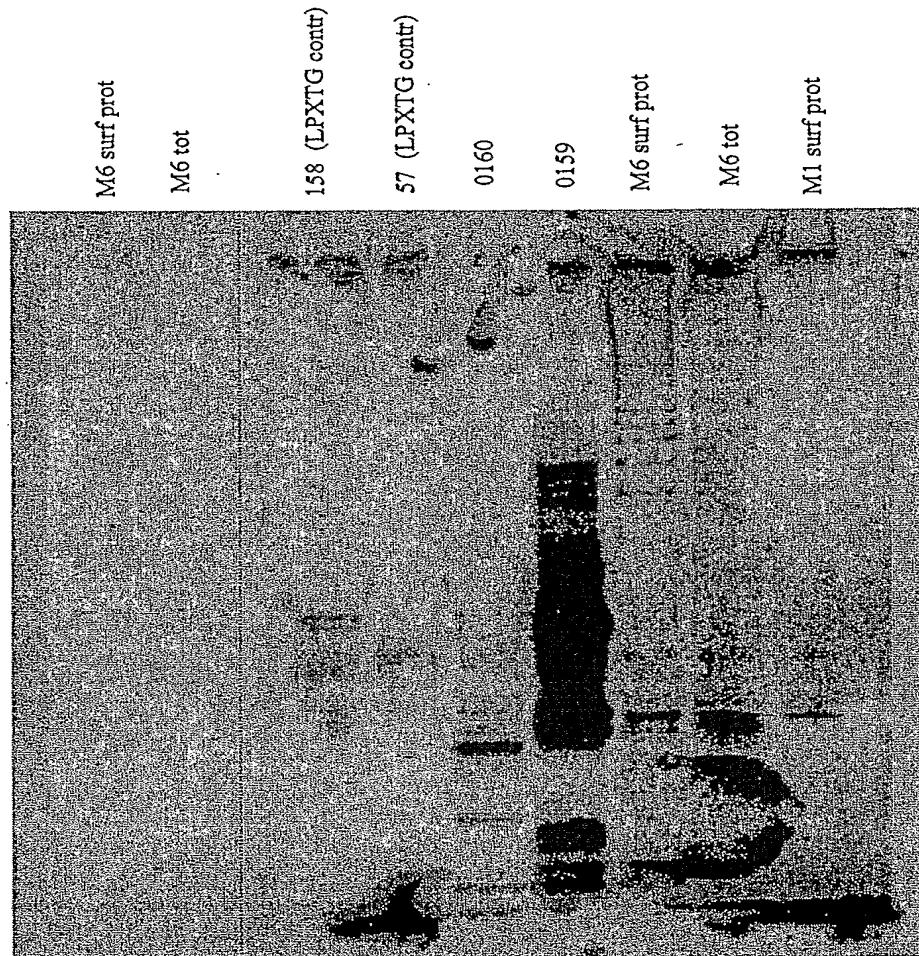
M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α -#: immune sera against #
P α -#: pre-immune sera

P α -18

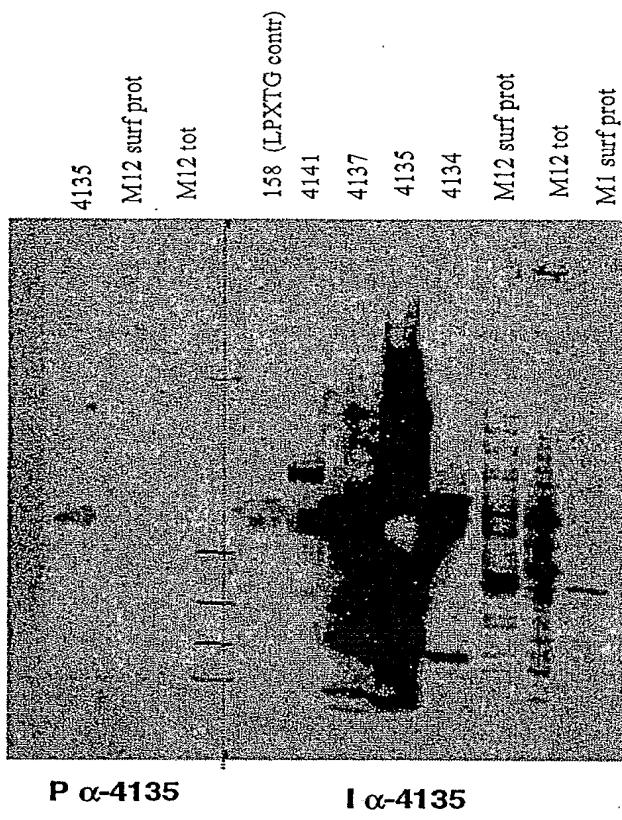
Figure 97

PCT/US2005/027239
248/487

Figure 98

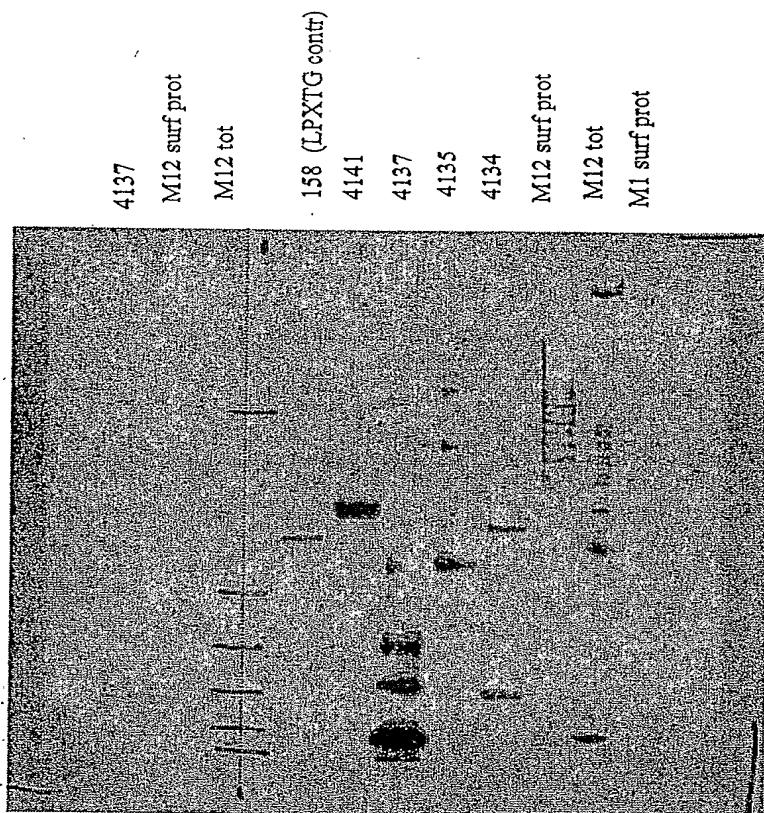
**LEGEND:**

- M6 tot: total extract (M6)
- M6 surf prot.: fraction enriched in surface proteins
- #: Purified recombinant proteins, 30 ng
- I α -#: immune sera against #
- P α -#: pre-immune sera

**LEGEND:**

M12 tot: total extract (M12)
M12 surf prot: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 99

PCT/US05/27239
250/487**LEGEND:**

M12 tot: total extract (M12)
M12 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 100

PCT/US05/E27239
251/487

FIGURE 101

GACAGGCTTCCTTATACGGACCGCTTTCTATAATGGACTCTTCCAAAGTTCTTGCTCTTTTA
 1 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 60
 CTGTGCAAGGAATATGCTGGCAGAAGATATAGCCTGAGAAGGTCAAGAACGAGAAAT

CCAGGGACTAGOCGTTCAGGTGCACAGATTGCTGGTGGTTGTTAAATGGAACCGAGTCGT
 61 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 120
 GGTOCCTGATCGCAAGTCCACGTTGCTAACAGCCACCAAAACAATTACCTGGTCAGCA

TCAAGTTGTGACAGAAATTACCTTCATCTGGATTCCGTTATGTTGGAGCTAGTGCC
 121 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 180
 AGTCAACACTGTCTTAAATGGAAAGATAGAACCTAAAGGCCAATACAAACCTCGATCACGG

TAAAAAGATTTCAAATTGTAAGCCGGGAAACTCTTGAGCTTGTGGCAATGTGTTTTTG
 181 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 240
 ATTTCCTAAAGTTAACACTTTCGGCCTCTTGAGAACTCGAACCTGAAACCGTTAACAAAAC

CTCTTGGTCGGATGGGAGTAGCTTGTGGGTCAAGCATGGCTATTGCTCTTGAC
 241 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 300
 GAGAACCGCGCTACCCCTCATCGAAAACGCCAGTGTGTTACCCGATAAGCGAAACTGG

AGCTATGTGAAAAAACACGACTTCACCCCTTTGGTAAATACCGTATCGTGTGGTAGT
 301 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 360
 TCGATAACACTTTTGTGCTGAAGTGGAAAACCATTATGGCATAGGCACGAACCATCA

GTTTGCTACTTTACAGTTTCGGTTTATTTGTATAAGAAAACCTTGAAGGGTAAC
 361 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 420
 CAAACGATGAAATGTCAAACAGGCCAATAAACATATTCTTTGGAAACTCCCCATTG

TCTTCAGGGTTTATACCTCTAGAAAATCTCTCAAAACGGCTAGCTTATCTGCAACC
 421 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 480
 AGAAAGTTCCAAGGAAATATGAGAATCTTTAGAGAAGTTGGCAGTCGAATAGACGTTGG

Figure 101A

TCAAAACAGTGTGTTGAGCAGCCCTGCCATAGTTCTCGCTTACGGATTCATT
 481 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 540
 AGTGTGTGTCACAAAACCTCGTCCCAGTGCAGGATCAAACGAGAACTAAAGTAA

GAGCTTTAAATCAGTCAGGGTAATCCCCAATAGGGGACACCTCTTCTTCGCTT
 541 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 600
 CTCGAATTGGCTCAGTCCATTAGGGTTATCCGGCTTGTTGAGGGAA

AATTCTTCATAGAGTTGCAAGGCTTATGGCTTATCTGACTAGCATCTTGTGTTTGG
 601 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 660
 TTAAGGAAGTATCTAACGTCCGATAAACGAATAGACTGATCGTAGAACCAAAACC

CAAGACTTTCGTTGGTAAGAGTTGAAAAGTCCTCTCGTAGGGATTCTAACATGACA
 661 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 720
 GTTCTGAAAAAGCAAACCTCTCAACTTTCAAGGACATCGCTAAAAGTTTACTGTT

TTTTCCAGTTTCTTGTGATGTTGAGTTGAGGGACTTTCTGATAGAGAGTCA
 721 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 780
 AAAAGGTGAAAGAACATCACATCTAACCTCTGCTGAAAGACTATCTCTCAGT

GCTCTTTGATATCTTCCCTGGCACGGAGAACCTCCCGTAGGGTTCTCTGCGGA
 781 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 840
 CGAGAAAAAAACTATAAGGGACCGCTGCTCTAGAAGGGCATCCAAAAGGGAACGGCT

TTGATTTACGGATTGGGATTGGATTGACTGGAGTTGTAATGCCACGGCTTCGAT
 841 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 900
 AACTAAATGCCTACGCTAACCTAAACTGACCTCTCAACACTTACGGTGGCTGGAAAAGCTA

ACAGATCATAGCCTAGTACCAAAACGGCTTACCTCAGGAACCTCAAGTA
 901 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 960
 TGTCAGTATCGGATCAGATGGTTTGCAGATAATCCAATGGAGTCCTGAGTTCA

253/487

Figure 101B

AATCAGGCCAGGTTAAAACGCCATTGGATGAAAGACGTTCTACTGTCTTCTACTC
961 TTAGTCGGTCAATTGGGTAAACTACTCTGCAAGATGACAGAAAAAGGATGAG 1020

CATGAATTTGAAATATCCATTGGTTGAGAAATCCTCAGCCGTTCAGGTAGAACATCA
1021 GTACTTTAACCTTTAGGTAAACAAACTCTTTAGGAGTCGGACAAGTCATCTTAGT 1080

CTGTCAAACCATTGGTTTGATAATCACTCGCCATTAGCTAAGAATTGTTGTAAG
1081 GACAGTTGGTACACCAAAACTATTAGTAGGGTAAATCGATTCTAAACAACATTC 1140

AAACGCCCTGGGGAGGAGCTTGAATGGGGTTCTTCCAGATACTCTTTGAATGGGGAG
1141 TTTGGGACGCCCTCGTCAAACTCACCTCAAGAAAGGCTATAGAAAAACTTACTCCGGCTC 1200

CAATTTGACCGCTGACTTGATAACCGAGTTTATTCGTACATCCAATAGGCTTCGT
1201 GTTAAACTGGGACTGAACATGGCTCAAATAAAAGACAGTGTAGGTTTATCCGAAGCA 1260

CAATGCTCATGGGTCAATCAAATCTGTTAGGCCTTAAAATAGCTCGAAATCCGGAGTC
1261 GTTACGAGTACCCAAAGTTAGGTTAGACATATCGCGAATTTCATGGCTTAGGCCCTCAG 1320

CCACAGACTTGTATTCTCATTAATTCCCTGAGATAAAGACAGCCCTGGGGACAAACGTTCAT
1321 GGTGGCTGAACATAAAGAGTATTAAAGGGACTCTATTCTGTCGGACCCCTGTGCAAGTA 1380

AAGCTTCCTGGAACTCATGGCAGAATGGACACCAAAGCTCTTGCCTCATAAACTACAGG
1381 TTGGAAGAACCTTGAGTACGTTACTCTGGTTTACCTGTTGAGAACGGAGTATTGATGTCC 1440

PCT/US05/27239
254/487

Figure 101C

1441 TAGAACGACTCCCCGTCCACCGTTGGCGAGGGTTCCTTCCAAATGACAGGTTTC
ATCTTGCTGAGGGCAGGGTGCACAAACGGCTCCCAGCAAGGTTATTACTGCCCCAAG
CTCTGAGTTAGGATTATCCCTGATTCCACTGCAGCAAAAAGGCATCCATGTCATAAT 1500

1561 GAGACTCAAATCCTAATGGGACTAAAGGTGACGTGTTCTCTAGGTACAGTTATA
GGATGATTTCCTTGACAAATCATTAAACAAAGGAAATCACATGCCCTAGCACCTTTT 1560

CCTACTAAAAGAACTGGTTAGTAATTGGTTCTCTTTAGTTGACGGATGGGGAAA
TATACTCTCGAAAATCTCTCAAAACCACGTCAAGCTTCCATCTGCAACCTCAAAACAGTA 1620

1621 ATATGAGAAGGCTTTAGAGAAAGTTGCTGAGCTGAAGGTAGACGGTAGGGACTTTGICAT
TTTTGAGCTGACTCGTCAGTTCTTACAACCTCAAGCAGTGCCTTGGAGTTGCGAAACTCGTGGACG 1680

1681 AAAACTCGGACTGAAGCAGTCAGATAAATGGGAGTTGCTCACGAAACTCGTGGACG
GGCTAGTTCCCTAGTTGCTTTCGATTTCATTGAGGTGAACTGGCTTATTTCTTTAT 1740

1741 CCGATCAAAGGATCAAACGAAAGCTAAAGGTAACTCACATGACGAAATAAGAAAATA
TATACCCCTTTCTGAAAAAAGAAAAAGGACTTTTATTTTTCAAAATAATAACA 1800

1801 ATATGGGAAAAAGACGTTTCTTCTTCTGAAATAAAAGTTTATATTATGT
GTTTGAATAATAAGACTGTTTGTAGAAAAGAAAGTGTAAAATAGGAATTTCAC 1860

1861 CAAACTTTATTTATCTGACAAAATCTTCTTCTGATTTTACATTTTTAAAGTGA 1920

Figure 101D

1921 TGTTGAAATCGGTTACTTTATGGTATACTTGTCTCATGAATGTAACAGATGACTGTTACT
ACAACTTTAGCCAATGAAATACCATATGAAACAGAGTACTTACATTGTCTACTGACAATGA

1980

1981 AGAAAAGGGACATTAATATGGTTGTTAAGACAGTTGTTGAAGGCACAAAGATAATTG
TCTTTTTCCTCGTAATTACCAACAAATTCTGTCAAACAAACTTGTGTCATATAAAC

2040 M V V K T V V E A Q D I F D - this orf is the homologue of sp0459, a formate acetyltransferase (pf1). It is out of the pilus locus

2041 ACAAGCTTGGAAAGGCTTCAAGGCCTAGATTGAAAGAAAAAGCAAGTGTATCACGCT

2100 TGTTTCGAAACCCCTCCGAAAGTTTCCGCAATCTAACCTTTCCTTTCGTTACATAGTGGCA

C K A W E G F K G V D W K E K A S V S R F -

2160 TTGTTACAAGCTTAACACACCTTATGATGGAGACGAAAGCTTCTCCTTGCAAGGACCAACAG AACATGTTGATGTTGAAACTACCTACCTCTGCTTTCGAAGGAACGTCCTGGTTGTC

C V Q A N Y T P Y D G D E S F L A G P T E -

2161 AGGGTTCACTTCACATCAAGAAATTTGAGAAGAAACTAAAGCACACTACGAAACTC TCGCAAGTGAAGCTGTTCTCTTCAACATCTCTTCAATTGTTGTTGATTTGCTGTTGAG

2220

C R S L H I K K I V E E T K A H Y E E T R -

2221 GTTTCCCAATGGACACTCGTCCAAACATCTATCGCTGATATCCCTGCTGGATTATCGACAA

CAAAGGGTTACCTGTGAGCAAGGGTTGAGATAAGCGACATATAGGGACCGACCTAAATAGCTGT

C F P M D T R P T S I A D I P A G F I D K -

2281 AAGAAAATGAAAGTTATCTTGGTATCCAAAATGAAACTCTCAAAATTGAAACTTCATGCT

TTCCTTTACCTCAATAGAAACCAATAGTTTACATACCTTACATGAAAGTTAACCTGAAAGTACG

2340

Figure 101E

C E N E V I F G I Q N D E L F K L N F M P -
 CAAAAGGTGGTATCCGTATGGCTGAAACTACTTTAAAGAAAAATGGATAACGAAACGCC
 2341 GTTTTCCACCATAGGCATACCGACTTTGATGAAATTTCCTTACCTTATGCTTGGCTGG + 2400

C K G G I R M A E T T L K E N G Y E P D P -
 CAGGTGTTACGAAATCTTCACTTAATAATGTAACAAACAGTTAACGACGGTATTTCGGTG
 2401 GTCGACAAGTGTCTTAGAAGTGAATTACATTTGTTCAATTGGCTGCCATAAAGGCAC + 2460

C A V H E I F T K Y V T T V N D G I F R A -
 CCTTACACTTCAAATATTCTGGTCCGCTCGTCATGCCACACTGTAACTGGTCTTCCAGATG
 2461 GGATGTGAAGTTATAAGCAGGGGAGGAGTACGTGTCGACATGACAGAGGTCTAC + 2520

C Y T S N I R R A R H A H T V T G L P D A -
 CATACTAACGGGAGCATCGGTGTTACGCACCTCTGCTCTTACGGTGCAGACT
 2521 GTATGAGTGGCCTGCAATGAGCCACAAATGCGTCCAGAACGAAATGCCACGTCTGA + 2580

C Y S R G R I I G V Y A R L A L Y G A D Y -
 ACTTGATGCCAAGAAAAGTAAACCACTGGATGCCAATCAAGAAATGCCATGAAACAA
 2581 TGAACTACGGTCTTTCATTGCTGACCTTACGTTAGTTCTTGTACTTCTTGT + 2640

C L M Q E K V N D W N A I K E I D E E T I -
 TCCGGCTTCTGGAAAGTAAACCTTCAATACCAAGCATGGCAACAGTTGCTGCTGG
 2641 AGGGAGAACCACTTCCTTCATTGGAAAGTTATGGTTCTTAACCTTGTCAACAGGGGACC + 2700

C R L R E E V N L Q Y Q A L Q Q V V R L G -
 GTGACCTTACGGGGTTGATGTTGGCTGCCAANCCAGCGATGAAACGTAAGCAATCCAAAT
 2701 + 2760

PCT/US05/27231 257/487

Figure 101E

C	CACTGGAAATGCCCAACTACAAGCCTTTCATGGCTGCTGGCGTGTGATTAAACGGTGTGCTGCTACATCTCTAG	
2761	GGCTTAACATTGCTTTCATGGCTGCTGGCGTGTGATTAAACGGTGTGCTGCTACATCTCTAG	+ 2820
C	CCCATTGTAACGAAAGCTTACCGACAGGGCACACTAATTGCCACGAGATGTAGAGATC	
V N I A F M A V C R V I N G A A T S L G -	GTCGTTGTAACCAATCGTATTGGACATCTTGCAGAACGTGACTCTTGCTCGGTTACATTAA	+ 2880
C	CAGCACATGGTTAGCATAACCTGTAGAAACGTTGACTGGAACGAGCACCAGTGTAAAT	
R V P I V L D I F A E R D L A R G T F T -	CTGAAATCAGGAATCCAAGAATTCTGTGATTGATGTTATGAAACTTCGTTACAGTTAAAT	+ 2940
C	GACTTAGCTCTTGTAGGTCTTAAAGCAACTACTAAAGCAATAACTTTGAAGGATGTCAATTAA	
E S E I Q E F V D D F V M K L R T V K F -	TTCGCTGTACCAAGCTTATGACCAATTGTACTCAGGTGACCCAACCTTTATCACAACCTT	+ 3000
C	AACAGGCATGGTTTCGAAATACTGGTTAACATGAGTGGTCCACTGGTTGGAAATAGTGTGTTGAA	
A R T K A Y D Q L Y S G D P T F I T T S -	CTATGGGTGTTATGGGTAAAGCAGGGTCTGTCACCGTGTACTAAGATGGACTTACCGTTTC	+ 3060
C	GATACCGACCATACCCATTGCTGCCAGCAACTGCGCAATGATTCTACCTGTATGGCAAAGA	
M A G M G N D G R H R V T K M D Y R F L -	TGAAACACTCTGACAACATGGTAACCTACCGGAAACCAAACTTGACAGTTCTTGGACTG	+ 3120
C	ACTTGTGAGAACACTGTGTAGCCATTGAGTGGTCTTGGTTGGAAATGTCAGAACCTGAC	
N T I D N I G N S S P E P N L T V L W T D -		

PC T/US05/E7E3258/487

Figure 101G

-ACAAATTGGCAACACTCCGTACCTGACATAGGCCAACAAACACTTCTTA
3121 -----+-----+-----+-----+-----+-----+-----+-----+
TGTAAACGGTATGTTGAGGCCAGCATACGTTTACTCGGTGTTGTGAGAGAT
3180 -----+-----+-----+-----+-----+-----+-----+-----+

C K L P Y N F R R Y C M H M S H K H S S I -
TICCAAATCGAAGGTGTAACACAATGGCTAAAGACGGATATGGTGAATGGCTGTATCT
3181 -----+-----+-----+-----+-----+-----+-----+-----+-----+
AGGTATATGCTTCCACATTGTTTACCGATTTCGCCTTACCACTTACTGACATAGA
3240 -----+-----+-----+-----+-----+-----+-----+-----+-----+

C Q Y E G V T T M A K D G Y G E M S C I S -
CATGCTGTGTGTCACCTTGATCCAGAAAATGAAAGAACACATCAGTACT
3241 -----+-----+-----+-----+-----+-----+-----+-----+-----+
GTACGACACAGAGGTGAAACTAGGTCTTTACTTCTGTGGGTGTAGTCATGA
3300 -----+-----+-----+-----+-----+-----+-----+-----+-----+

C C V S P L D P E N E E Q R H N I Q Y F -
TGGGTCTGTAAACGTTCTTAAAGCCCTCTTACTGGTTGATGGGTACGACG
3301 -----+-----+-----+-----+-----+-----+-----+-----+-----+
AGCCAGGACACATTGGCAAGAAATTGGAGAATGACCAAACTTACCAATGCTGC
3360 -----+-----+-----+-----+-----+-----+-----+-----+-----+

C G A R V N V L K A L L T G L N G G Y D D -
ATGTTACAAAGACTACAAAGTTGATATGAACTTCCGTGACGAAGTTCTTGAT
3361 -----+-----+-----+-----+-----+-----+-----+-----+-----+
TACAAGTGTCTGATGTTCTATAAAACTATAGCTTGGCTAGCTGCTTCAGAACTTA
3420 -----+-----+-----+-----+-----+-----+-----+-----+-----+

C V H K D Y K V F D I E P I R D E V L E F -
TTGAATCAGTTAAGCGAACTTGAAAAATCTCTTGACTGGTTGACTGACACTACGTAG
3421 -----+-----+-----+-----+-----+-----+-----+-----+-----+
AACTTAGTCATTTCGCTTGAACATTAGAGAACACTGACCAACTGACTGTAATGCATC
3480 -----+-----+-----+-----+-----+-----+-----+-----+-----+

C E S V K A N F E K S L D W L T D T Y V D -
ATGCCCTTGAACATCATCCACTACATGACTGATAGGTACAACACTACGAAAGCTGTTCAAATGG
3481 -----+-----+-----+-----+-----+-----+-----+-----+-----+
TACGGAAACTTGGTAGGTGACTGACTATCCATGTTGACTGCTTCGACAAGTTTAC
3540 -----+-----+-----+-----+-----+-----+-----+-----+-----+

PCT/US05/16505 / 27239
259/487

Figure 101H

c A L N I I H Y M T D R Y N Y E A V Q M A -
 CCTTCCTGCCAACTAAACAACGTGCGAACATGGATTCTGGTATCTGGATTGGCTAACAA
 3541 GGAAGAACGGTTGATTGTCACGGTTGACCTTAAGCCATAGAACCTAACGGATTGT 3600

c F L P T K Q R A N M G F G I C G F A N T -
 CTGGTGTACATTTGTCAGCTATCAATAAGCTACAGTTAACCAATCCGTGACGAAGATG
 3601 GACAACATATGTAACAGTCGATAGTTATGGCATGGTCAATTGGTAGGCACITGGTTCTAC 3660

c V D T L S A I K Y A T V K P I R D E D G -
 GCTACATCTACGATTACGAAACAAATCGGTGACTACCCACGTGGGTGAAGATGACCCAC
 3661 CGATGGTAGATGCTAATGCTTTGGTTAGCCACTGATGGGTGGACCCACTTCTACTGGGTG 3720

c Y I Y D Y E T I G D Y P R W G E D D P R -
 GTTCAACGAATTGGCAGAATGGTTGATCGAAGCTTACACAACACTGCTTACGTAGCCACA
 3721 CAAGTTGCTTAACCGTCTTACCAACTAGCTTCGAAATGTTGAGCAGATGCTCGGTGT 3780

c S N E L A E W L I E A Y T T R L R S H K -
 AACTATACAAAGACGCCAGCTACAGTATCACTTTGACAATCACATCTAACGGTTGCTT
 3781 TTGATATGTTCTGGCTCTTGATGTCATAGTGAACACTGTTAGTGTAGATTGCAACGAA 3840

c L Y K D A E A T V S L L T I T S N V A Y -
 ACTCTAAACAAACTGGTAACCTACCAAGTTACAAAGGGTATAACCTCAACGAAGATGGTT
 3841 TGAGATTGTTGACCATGGTGGTCAAGTGTTCCACATATGGAGTTGCTTCTACCAA 3900

c S K Q T G N S P V H K G V Y L N E D G S -
 CTGGTGAACTTGTTCTAAACTTGAAATTCTACCCAGGTGCTAACCATCTAACAAAAGCTA
 3901 3960

PCT/US05/27239 260/487

Figure 101I

GACACTGAAAGATTGAACTTAAGAAGGTGGTCCACGAATGGTAGATTGTTTCGAT
 C V N L S K L E F F S P G A N P S N K A K -
 AAGGTGGTGGTGCAAAACCTTGAACTCACTTCTAGCCTTGACTTTAGTTATGCCAGCTG
 3961 -----+-----+-----+-----+-----+-----+-----+-----+ 4020
 TTCCACCAACCAACGTTTGAACTTGAGTGAAGATGGAACTGAATCAAATGTCGAC

G G W L Q N L N S L S L D F S Y A A D -
 ACGGTTATCTCATGACTACACAGPATCACCTCGGGCTCTGGTAAGACTCGTGTGATGAAAC
 4021 -----+-----+-----+-----+-----+-----+-----+-----+ 4080
 TGCCTATAGACTGAACTGATGTTCTCATGTTGAGCGGAGAACCATCTGAGCACTACTTG

G I S L T T Q V S P R A L G K T R D E Q -
 AAGTTGATAACTTGTAAACATCCTTGATGGTTACTTGAAAACGGGGACAAACAGTTA
 4081 -----+-----+-----+-----+-----+-----+-----+-----+ 4140
 TTCAACTATTGAACTTGTAGGAACTACCAATGAGGCTTTGCACCTGTGTGCAAT

V D N L V T I L D G Y F E N G Q H V N -
 ACTTGAACTTATGGACTTGAACTGTTACGAAAAAAATCATGTCAGGGAAAGACGTTA
 4141 -----+-----+-----+-----+-----+-----+-----+-----+ 4200
 TGAACCTTGCAATAACCTGAACTTGCTACAAATGCTTTTAGGAGTGAAGTCTGTGCAAT

L N V M D L N D V Y E K I M S G E D V I -
 TCGTACGTATCTCTGGATACTGTTAAACACTAAATACTCCTACTCAGAACAAAAACTG
 4201 -----+-----+-----+-----+-----+-----+-----+-----+ 4260
 AGCATGCAATAGAGACCTATGACACATTGTTGATTAGGAGTGAAGTCTGTGTGAC

V R I S G Y C V N T K Y L T P E Q K T E -
 AATTGACACAAACGGTGTCTCCACGAAGTTCTCAATGGATGACGCCCTGGATGCAATTGA
 4261 -----+-----+-----+-----+-----+-----+-----+-----+ 4320
 TTAACCTGTTGACAGAAGGTGCTTCAGAAAGTTACCTACTGGAAACCTAACGTAACT

L T Q R V F H E V L S M D D A L D A L S -

PCT/US2005/223333 filed

261/487

Figure 101J

	GCTAATCAAGTTCTTGGATAATAAAAGGGCTCTTTCGACTGTAGGGTTGAGAAA		
4321	CGATTAGTCAAGAACTTATTATTTCGGAAAACAGTGCACATACCCAACTTCTTT		4380
*			
	AGCTTAAGCTCGAGAAAGGACAATTTCGCTTTCCTTGTGTTTGATGTTCAAGGGATGAAA		
4381	TCGATTGAGCTCTTCTGTAAACAGGAAGAAAAACTACAAGTCTGCTACTTCTT	4440	
e	* A R S L V F K T R E K K I N L A I F -orf1_670 homologue of sp0460, transposase		
	ATCCGTTTTTGAAAGTTCAAGGTTCCGAAACCAAGGCATTGGCCTTGATGTTCTTGG		
4441	TAGGCAAAAACCTCAAAAGTTCAAGGCTTTCGGTTCCGTAACGGAACTACAGAAAC	4500	
e	I R K K F N E F N R F G F A N R K I D K -		
	ATGAGTTTGTAGTGGCCTCAAGTTAGCGTTAGAATAAGGCAATTCAATGGCTTAGTG		
4501	TACTCAAACCAATTCAACGGGAGTTCAATCGCAATTCTTATTCCGTTAACGTACCGCAATCAC	4560	
e	I L K N T A E L K A N S Y P L E I A N T -		
	ATGAGTTTATAGCAAAATAATGCTCAAGTGGTTAAAGGTGGTTGAGATGA		
4561	TACATCAAAAATATCGTTTATTACACGAGTTCAACGAGTTCAACAAATTCCACGCCAACTCTACT	4620	
e	I Y N K Y C I F T S L T T K F T R N L H -		
	GGTAACGGTCTTGTAAATTAGCCCCAAAACCTGGTCAGTATCTCTCTGTAGATGAAT		
4621	CCATTGCAAGAACTTAAATCGGGTTTGACCGTCAATAAGAGAACATCTACTTTA	4680	
e	P L T D Q I L G W F Q D T N K E Q L H F -		
	AGGAGTAGTTGATAACGGTCATAAGTAATCTTAAAGTTCAAGTCTAGGTACTAGAGTAAAGATTTTC		
4681	TCCCTCATCAACTATGGTCCAGTCAATTAGAAATTCAAGTCATGATCTCATTTCTAAAG	4740	

PCT/US605/27239

262/487

Figure 101K

e L I L Q Y I D Y D K E P V L T F I K -
 4741 TTCAGACACTCCCTAGGAGTTAAGGTCTCTGAAGGTTCTAGCATAGAAAGGCTTAAGA
 AAGTCGTGAGGGATCCTCAATTCCAGAGGACTTCAAGATCGTATCTTCCGAATTCT + 4800

e K L C E R P T L T S R F T R A Y F P K L -
 GAGAGTTCCGACTATCCTTTAGGATAAATTTCAGTAAATTAAAGGCTCTGTTATTCC
 4801 CTCTCAAAGGCTGATAGAAAATTCTATTAAAGGTCAATTAAATTCTCGAGAACATAGG + 4860

e S L K R S D K L I F K W Y Y K L A R Y E -
 AGAGATTATCATCAAATTGCTTCATGATGTTGATTCTAGTCGATTAAGAGCCCTGCAC
 4861 TCTCTAAATTAGTAGTTAACGGAGTACTAACAACTAACGACTAAAGATCAGACTAAATTCTCGGGAG + 4920

e L S K D D F Q K M I N I R T Q N L A R S -
 ATGGTTGGACAATGGGAAACGGATCGAGAACAAATTAGGATTGGAAATTAAATTCTTA
 4921 TACACAACCTGTTACACCTTGTCTTGCTAGCTCTTGTTAAATTCTGAACCCATTAAAGAAT + 4980

e M H Q V I H F R D L V I K A N P F L K K -
 ATGAGGGATAACTTCCAGACATACTAACAGTGACGACTTTAACCTTTCTAGCT
 4981 TACTCTCCCTATATTGAAGGTCTGTTAGGTGTCAGCTGCAATTGAAAAAGATGCA + 5040

e I L P I Y S G S M D V T V V K V K R A -
 TCTTTGAGTACTTGAAGAAATGATTGGATGGTTGGCTGTTACAGAACTGGAGAACATAGTCTTAC
 5041 AGAAAGCTCATGAACTTCTTAACTTAAAGCTAACCAACACTGGAGAACATAGTCTTAC + 5100

e E K S Y K F F H N R I T T Q R R N D L I -
 GTCTAGATTTCTAGCTTAACTGAGGAAATGAAACCCATTCCCTCTGGTAG
 5101 + 5160

263/487

Figure 1011

CAGTACTAAAAGAATCACACCTTAAGGACTCGTTACTTTCGGTTAAAGGGAAAGACCATC

e T M I K K T N F D Q A I F A L K G K Q Y -

GAGAATTTCATCCCAGGAGGGATTTCAGGCAAGTGGTGTAAATCCTCTTGGAAATGAAT 5161

CTCTTAAGTAGGGTCCCTCTCTAAAGTCCGGTTAACACATTAGGAGAACCTTACTTTA

e S F E D W S L I E P L T T Y D E Q F H F -

TGCCTTGGCTTACCGATAAGCCTAGACCTAGAGCTAGAGCATGGCTTAATTAGAACGGATA 5220

ACGAAACTCGAAATGCTATCTGCATCTCCATCTCCATCTACCGATTAATCTTCGCTAT

e Q K L K R Y V T S T S T S I A L K S A I -

TGTGTAAGGCCCTCTCTGTAGTAGGAGGTGGCAATTTCAGCTACAGTGACTTTCGACAGGACTTG 5281

ACACATTCTGGAGACAACATCCATCCCTAACCGGTAAAGACAGAGTGGTAAGGGCTC 5340

e H T L A E R N L L Q A I K Q R V M E S -

ATTTGGCAATTTCAGCTACAGTGACTTTCAGCTACAGTGACTTTCGACAGGACTTG 5341

TAAACCGTTAAAAGACTTGCTCTCAACAAAGTCGATGTCACTGAAGGCTGTCCTGAAC 5400

e I Q C N K Q V L T F E A V T V K R C S K -

CATTGAAATCGTCCTCTTTCAAATGAATGAGGCTAGGGAAACCAATCTCGATAAAA 5401

GTAACCTTGTAGAGAAAAAGTTACTTACTTCGATCCCTGGGGTTAGGGTATTT 5460

e C Q F R R K L H I L S P F G G I E I F -

GGGATTTTAGAAGGCTTTGGAAAGTCGTATTGATTGTTTCCTTACAGTGTTCAT 5461

CCCTAAATCTCGAAAACCTTCAGCATAAACTAAACAAAAGGAATGTCACAAATGTA 5520

e P I K S P K Q F D Y K I Q K C H K C -

PCT/US05/27239 264/487

Figure 101M

TTAGGGTGGGTGATAATCAAGTGTAGCGAAGACTTCGATATGGGTATCGTGCTGAATGGCT
 5521 +-----+-----+-----+-----+-----+-----+-----+-----+
 AATCCACCCACTATTAGTTACATCGGTTCTGAAGCTATACCCATAGCACGACTTACCGA
 e K P P H Y D L T A F V E I H T D H Q I A -

 TTATTAAAGGTGATGTTTGTCTTTATTCCGATGAGTAATGGGTATCAGTGTGTT
 5581 +-----+-----+-----+-----+-----+-----+-----+-----+
 AATAAATTCCACCTACAAAACAGAAAATAAGGCTACTCTATTACCCATACTAACACA
 e K N .L T I N K D K I G I L L T T H N I H -

 TCCATAAGATACTTCTAAATGAGTTGTTAGGGCTTTTCATTATAAGCTTATGGACT
 5641 +-----+-----+-----+-----+-----+-----+-----+-----+
 AGGTATTCTPATGAAAGATTACTCAACAAATCCGGAAAAGTAAATTAGAATACCCCTGA
 e E M -

 TTTTTGAACTCAAAAAGCCCTATAATTCTCCACAGGGATTTACCCACTACAGAAATT
 5701 +-----+-----+-----+-----+-----+-----+-----+-----+
 AAAACTATGAGTTTTCGGGATTATTAGAGGTGTCAACCCCTAAATGGGTATGCTTTAAT

 TAGAGCCAGAAAAAACACTTTTGTTCACCTAGCAGAAACTAGAGGCCAAAGTGTTTCT
 5761 +-----+-----+-----+-----+-----+-----+-----+-----+
 ATCTCGGTCTTTTGTGAAACAAAGTGTCTTGTATCTCTCGTCTTCACAAAAAGA

 GTTCAGATTACCCAAAACTGGGAAATAATGGGTATAGAATAGAGATGGCTTAGGAAGCC
 5821 +-----+-----+-----+-----+-----+-----+-----+-----+
 CAAGCTTAAATGGGTTTGCCTTTTAAACCCCTATTCTPATCTTACCGAAATCTCGG
 e M -

 CCTTTTGTAGACAGTACGATGAACTTACAAATACTGAGCTTATTCACCTGGAAATC
 5881 +-----+-----+-----+-----+-----+-----+-----+-----+
 GGAAAAACACATCTGCTCATCTGCTTACCTGGAAATAATCGTGTGATAGATGAGA
 e * L L Y H A K K A I -orf2_670 homologue of sp0461, transcriptional
 ATTGGGACCCGTTTGTCAAAGCCTCTTTTGGATATCTACATTGCTGATAGATGAGA
 regulator

265/487

Figure 101N

5941 TAACGGCTGGCAAACAGTTTCGGAGAAAGCCATATAGATGTTAACAGACTATCTACTCT 6000

e M A V R K D F A E K R I D V I T Q Y I L -

CGCTGTGGTACATGCCAATCTAAGGCAATCGTCAAAAAGTGAATGTTCCCTTTGGGA,
6001 GCGACIACCGGAACTGACGTTAGATTCCGGTAGCAGTTTCACTACAAGGAAACCT 6060

e R Q Q S V H L D L A I T L F H H K G K P -

TACTGCCTTAAACGTAAGGCCAGGTTATTCTTCTGTTGTAATAATAATCAATGCGCTCTGTC
6061 ATGACGAAAAATTGCATTCCGTCCATAAGAAACATTATTAGTTACCGAGACAG 6120

e Y Q K K V Y P L Y E K T T I I L P E T -

AAATGCTCCTCTGAAGGGAGGACTAATTAGAATAATTGTATCCTGTAACAGAGGCACT
6121 TTTACGAGGAGACTCCCTCCCTGATAATCTTATAACATAGGACATTGTCTCCGTGTA 6180

e L H E E S P P P S I L I N Y G T V S A V -

TTTGTCAAGTAAATTCCGTAATAATGGACTTTATAAGTTAACATCTGCTTGATTATT
6181 AACAGTCATTAAAGGCATTATTACCTGAAATAATTCAAAATTGAGCAACTAAATAAA 6240

e K D T F N R L I I S . K I L N V D A Q N N -

AAATGATAAAATCGGGATAGCAGGTAGTGGAGAAAGATGGTTTCTGTCAAAGTAGAGT
6241 TTTTACTATTAGCCCTATGGCCATCACTCCCTTCTACCAAAAGACAGTTCATCTCA 6300

e L I I F I P I A P L S S F I T E T L Y I -

GAGGAAAGGTAAGCCGATGGCTGGTGGTCAACTCCCTCAATCTCTGCTCAAGTCATCCAC
6301 CTCTTTCCATGCGCTAGGACCAGCTATTGAGGAAGTTAGAAGCAGTCAGTAGGTG 6360

e S F L Y L R H Q D I V G E I K Q E T M , W -

266/487

Figure 1010

TCTTGAAACATTGCTTTCGAAATATGATAACAGTGGCTTGTGCTTCAATCCATAATGT
 6361 AGAACATTGTTAACGAAAGCTTATACTATGTCAACGGAACAGCAGAAAGTTAGGGTATTACA
 e E Q V I A K S I H Y L P K D S E I G Y H -

 TCGTAATAATTATAATAGGAACTAGATTTGTAACCAAACAAAACGTTCTTGTAAAG
 6421 AGCATTAATTATAATTATCCCTTGATCTAAACATTGGTTGGTTGGCAAGAACATTG
 e E Y Y N Y P V L N Q L G F L F T R T L -

 AAAGTCAGTGCTGTTAAAAAGAAAGAGAATTGCAAAATGTCATTCTTAAGATAATTCTTG
 6481 TTTCACTCACGACAATTTCCTCTTCAAGCTTACAGTAAGGATTCTATAAGAAC
 e F T L A T L F S L S N S I D N G L I N K -

 AACTTGGAATAGTAGATGCTTCTTGTATGCTGAAGAATCAGTTGAATAGGTATGAGTC
 6541 TTGAAACCTATCATCTAGAAAGGAAACATAAGCATTCTTAGTCACCTTACACTCAG
 e F K S L L H K G R T H Q L I L Q I T H T -

 TTTTTTCTGATTCCATTGCTCTGGAAACGAAAGAATTAGCAGAACATAACCAAA
 6601 AAAAGAAACTAAAGTAAACAGGAACCCCTTGCCTCTTAATCGTCCTGGTT
 e K K E Q N W K D K S F S S N A S C Y V L -

 AAGATATAATCCAGTTCTCTGAGTAAAGTCATGGGGCATGGCTCTAAGTAAGTT
 6661 TTCTATATTAGGTCAAGAAGGACTCATTTCAACCGTACACCGAGATTCATCAA
 e F I Y D L E E Q T F T M N A H P E L Y T -

 TGGCAATGTTCCATCAAATGGATACTAAAGGGTTTTAACTTTCAAAACTCTTTG
 6721 ACCGGTACAAGGGTAGTTTGGCTATGTATTCTCCAAAAAAATTAAAAGTTGAGAAC

Figure 101P

e Q C H E M L I P Y M F L N K L K E F E K -
 GACTCAGGAACTCAAGTGGAAATTCCCGACGTTCACAGTGGCCACTAGTAGATGCTA
 6781 CTGAGTCGCCCTTGAGTTCACCTTTAACGGCTGCACAGGGTTCACTCACGGTGAATCATCGAT

e S E P F E L P F E R R K W T L A V L I S -
 AAATGAAACATACTCGTCAGGGTGTGATTTCATAACAGTTCATGACTGAGTTGAGATTAGAC
 6841 TTTTACTTGTATGAGCAGTCCACACTAAAGATTGTCAAGTACTGACTCAACTCTTAATCTG

e F H V Y E D P T I E L L E H S L Q S N S -
 TGCACAAATCATATGTTGACCCAATCCATACTTCCATCAATTCAAATCTCAATA
 6901 ACGTGTAGTAACTACACTGGGTAGGTATGAAGGTAGTAAGTTAGTTAGACTTAT

e Q V I M H T V W D M S G D N L D Y I E I -
 CCAAAATGAAACTGGAGGTGCAATTAAAAAACGAAATGCGATATTAGGGCCAACTACT
 6961 GGTTTTACTTTGACCTCCCTACGGTTAAATTGCTTACGCTTAAAGTCCTGGTTGATGA

e G F H F Q L L A I L F R I R Y E P G V V -
 TGATTTTCAAGGTCCAACCTACTGAACGTAGTAACAAGCCACACTTTTGCGTAGG
 7021 ACTAAAAAGTGTCCAGGTTGGATGACTGCTCATGTTCCGGTGAACAGCATGC

e Q N K V L D L G V S R L L G C K Q R V -
 CGGTAGCCCTGTTGGATGGAAATAATACCTTTTGTGTAATTGCTTAAAGCTTGTGATTA
 7081 GCCATGGACAAACGCTACCTTTATGAGAAAACACATTAAAGCAATTGAAACTAAT

e R Y G T A I S I Y E K Q T F E N F S Q N -
 CCTTGATGAGAAAGGGAGTATTAAAATAGTTGATTTGATGAAAGCTGATGG
 7141

PCT/US05/27239

268/487

Figure 101Q

GGAACATCATCTTTCGCGCTCATAAAAATTTCACCAATAACCAATATTTCGACTACC
 e G Q L L F F R L I K L I T S Q N Y L Q H -
 AAGTAATAATTCTGGTTGATGAGAATGGGTGCGATTAAATTGAACCTGGTGGTATCTAAA
 7201 TTCAATTAAAGCAAACCTTACCTTACCAAGCTTAATTAACTTGAACAACGCATAGATT
 e F Y N T Q H S H E I L Q V Q Q T D L -
 TTAAATGTCAACTCTCCCTGAATGGTTCTGTAAATTCTGCAAATGCTTAGGAGACTT
 7261 AATTACAGTTGAGAAGGAGCTTACAAGAACATTAAAGGACGTTTACGAATCCTCTGAA
 e N F T E E E F T E Q L E Q L I S L L S -
 TTAGATTGTAATGGAGTTAAAGTAGACAGTTCATCTGTTCAATTAGACCGAAATATCCAAT
 7321 AATCTAACATTACTTCATCTGTCAAGTAGATCAAGTTATCTGGCTTATAGGTTA
 e K S Q L S T L E D L E I S R I D L -
 AATATATTAAAATGGTAATTTCATCTGTAATTCTTTCAATTGTTAGCTATA
 7381 TTATATAATTACATTAAATAGACATTAAAGAAAGTTACATAAACAAATCGTAT
 e L I N L I T I K D T I R K E I Y K N L M -
 GTTACCGAATCTTAGTTGCAATTAGATAATTTCATTAAATACAAAAGAAACTAAT
 7441 CAATGGCTTGAATCAACGTTATCTTAAATTAAATTGTTCTTGTTGTTGTTGTTA
 e TGTCTGTCAAAAAGGTTGTGGAATTCCGACTTTATGATAAAACAGCATGTAATAAA
 7501 ACAGAACAGTTTCAACACTTAAAGGTGAAATAACTTTTCGTCGTCATTATTT
 7560 TGGCATTTAAAGATAGTAATGGATTGGGGAGTTTATGGCTTATTTTATTAGA
 7561

Figure 101R

CCGTAAATTCTATCAATTACTATAACCACTTCAAAATAACCGAATTAAAAATAACT

AAATAATTCTTATCAAAATAATGCGTTCTATAAAAATAATGGATAAAAATACTATT
7621 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 7680

GTTATAAAAATAAGTTTATAACGCAAGATAATTCTATAGATAAA
7681 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 7740

ATTCGATACAAGTGTGTCATTGCCAGGTGAGAAGATAAGCTATAACGCACTTTATACGC
7741 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 7800

TAAACATATGTTAACAGTAAACGGTCCAACCTCTCTCATCGATAATTGCGAAAAATATGCG
7801 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 7860

TTTTCGTAACGTTTAGTGAACGGATTAACTCACTGTGAGATAAAATTATCAGAACATAA
7861 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 7920

AAAACGATGCAAACAAATCACITGCCATTAAUTGAGTCACUTCACTTAATAATGCTTGATT
7921 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 7980

TGTTCTATGAAATTGCTTAACAGGGAGAACACACATGAAAAGCTAGAAAGATAATTCA
7981 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 8040

b M L N R E T H M K V R K I F Q -orf3_670 homologue of sp0462, LPXTG
K A V A G L C C I S Q L T A F S S I V A -

PCT/US05/23239

270/487

Figure 101s

b TTTAGCAGAACCCCTGAAACCACTCCAGGGATAGGAAAAGTAGTGTAAAGGAGAAGG
 8041 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 AAATCGTCCTTGCGGACTTTGGTCAGGTGCTATCCTTTCATCACATAATTCCCTGTCC
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 b L A E T P E T S P A I G K V V I K E T G -
 CGAAGGGAGGCCCTCTAGGAGATGCCGCTTTGAGTTGAAACAAATACGGATGGCAC
 8101 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GCTTCCTCCTCGCGAAGATCCTCTACGGCAGAAACTCAACTTTTGTATGCCCTACCGTG
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 b E G G A L L G D A V F E L K N N T D G T -
 AACTGGTTTCGCAAAGGACAGAGGGCGAACACAGGAGAAAGCGATATTTCACATAAAC
 8161 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 TTGACAAAGGGTTTCCTGTCTCCGCGTTTCACCGCTCTTCGCTATAAAAGTTTGTATTTGG
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 b T V S Q R T E A Q T G E A I F S N I K P -
 TGGGACATACACCTTGACAGAACCTCCAGTTGTTATAAACCCCTCTACTAAACA
 8221 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 ACCCTGTATGTGAACTGTCTTCGGTTGAGGTCAACCAATAATTGGAGATGATTGT
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 b G T Y T L T E A Q P P V G Y K P S T K Q -
 ATGGACTGTTGAAGTTGAGAAGAAATGGTGGGACGACTGTTCAAGGGAAACAGGTAGAAA
 8281 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 TACCTGACAAACTTCACITCTTACCGCTGCTGACAGGGTCCACTTGTCCATCTTT
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 b W T V E V E K N G R T T V Q G E Q V E N -
 TCGAGAAGAGGCTCTATCTGACCACTATCCAGTAAACAGGGACTTATCCAGATGGTCAAAAC
 8341 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 AGCTCTTCCTCGGAGATAGCTGGTCATGGTGTTCCTGAATAGGTCTACAAAGTTG
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 b R E E A L S D Q Y P Q T G T Y P D V Q T -
 ACCTTATCAGATTAAAGGTAGATGGTGGAAAAAACGGACAGGCACAAGGGTTGAA
 8401 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 TGGAAATAGTCATAATTAAATTCATCTACCAAGGCCCTGGTGTGTTCCGCAACTT
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

PCT/US05/22323-9

6. *Am. J. Phys. Chem.*, 1886, p. 103; *ibid.*, 1887, p. 103.

271/487

Figure 101T

b	P Y Q I I K . V . D G S E K N G Q H K A L N -	
8461	TCCGAATCCATATGAACTGGTGTATTCCAGAAGGTACACTTCAAAAGAGATTATCAGT AGGCCTTAGTTACTTGACACTAAGGTCTTCATGTAAAGTTCCTTAATAAGTTCA	+ 8520
b	P N P Y E R V I P E G T L S K R I Y Q V -	
8521	GAATAATTGGATGATAACCAATATGGAATCGAGGTAGCGTTAACGGTAAACGACGGT CTTATTAACCTACTATTGGTTAACCTTAAGCTAACACTGCCAATCACCATTGTGCCA	+ 8580
b	N N L D D N Q Y G I E L T V S G K T T V -	
8581	TGAAACGAAAGAAGGCTCTACTCCGGCTAGATGTTGTTATTAGATACTCCAAATA ACTTTGCTTTCTTCTGGAGATGGGGATCTACAACTAGATAATCTTGGGTATTC	+ 8640
b	E T K E A S T P L D V V I L L D N S N S -	
8641	TATGAGTAATTTCGACATAATCATGCCATTGAGCGGGAAAAGGGGAGAAGGGACACG ATACTCAATTAAAGCTGTATTAGTACGGGTAGCTCGCCCTTTCGCTTGTGCG	+ 8700
b	M S N I R H N H A H R A E K A ' G E A T R -	
8701	AGCCCTTGTAGATTAAGATTAACCTCCATTCCAGAAATCGAGTAGCTTGTGACTTGG CTCAACTATCTTGTACGGTTCAAGAAGCTACTGTGGAAAAGGGTAGCAGATGGAAACGG	+ 8760
b	A L V D K . I T S N P D N R V A L V T Y G -	
8761	TGGGAACATCTATTCTTATGGAGCTTAAGCTTAATGGCTCACTCGTGAACACTGAATCC GAGTTGATAGAAAATGACTGCCAACTGCTTCGATGACACCTTTTCCCCATCGTCTACGGCTTGGCC	+ 8820
b	S T . I F D G S E A T V E K G V A D A N G -	
8821	AAAATATTGAAATGACTCAAGCTTTATGGACCTGATCGACGACTTACAGCTAAAC	+ 8880

Figure 101U

TTTTTATAACTTACTGAGTCGAAATAACCTGCAAGCTAGCATGGCTGCAAATGCTGATTG
b K I L N D S A L W T F D R T T F T A K T -
TTATAATTAGCTTTTTAATCTCACATCAGATCCTACTGATAATTCAAACTTAAAGGA
8881 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 8940
AATATTAAATATCGAAAAATTAGAGTGTAGTCTAGGATGACTTAAGTTGATAATTCT
b Y N Y S F L N L T S D P T D I Q T I K D -
TAGGATTCCATCAGATGCAGGAAATTGAACAAGACAATTGATGATAATTGGGGCG
8941 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 9000
ATCCCTAACGGTAGCTACGTCTCCCTTAACCTGTTCTGTTAACTACATACTAGTTAGGCCGCG
b R I P S D A E L N K D X L M Y Q F G A -
GACTTTACCCAGAAGGCTTGTGACCGCTGATGATACTCTGACAAAGCAGGAAAGACC
9001 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 9060
CTGAAAATGGGTCTTCCGAAACTACTGGCGACTACTATAGAACTGTTCTGTCGGTTCTGG
b T F T Q K A L M T A D D I L T K Q A R P -
AACAGTAAAAAGGTTATTTCACATTACAGATGGTGTCCGACTATGTCATATTCCAAT
9061 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 9120
TTTGTCATTTTCAATAAAAGGTGTAATGTCACCAAGGCTGATACTAGGTTA
b N S K K V I F H I T D G V P T M S Y P I -
TAATTAAATACAGGAACGACGCAATCGTACAGAACTCAGCTGATAATTAAAGC
9121 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 9180
ATTAAAATTTATGTCCTGCTGCGTTAGCATGTCCTGACTGACATTATAAAATTTCG
b N F K Y T G T T Q S Y R T Q L N N F K A -
AAAAACTCCAATAGTACGGGGATAATTACTGGGGACTTGTACATGGTACGGAGATGG
9181 +-----+-----+-----+-----+-----+-----+-----+-----+-----+ 9240
TTTTGAGGTTTATCATGCCCTATAATGACCTCCTGAAACATGGTACAGTGTCTAACC
b K T P N S S G I L L E D F V T W S A D G -

PCT/US05/27239

273/487

Figure 101V

b TGAACATAAGATTGTTGGAGATGGAAAGTTATCAGATGTTACGAAGAACCTGT
9241 ACTTGTATTCTAACAGAACCTTCAACCCTTCAATTAGTCCTACAATGCTTCTTGGACA + 9300

b E H K I V R G D G E S Y Q M F T K K P V -
9301 AACAGACCAATAACGGAGTTCATCAAATACTTCAATCACCTCCATGAGCAGAGCTAA + 9360
TTGTCTGGTTATGCCTCAAGTAGTTATGAAAGTTAGTGGAGGTACCTCGTCTCGATT

b T D Q Y G V H Q I L S I T S M E Q R A K -
9361 ATTAGTTTCAGGGGATATAGGTTCTATGGAACTGACTGTATTATATTGGGTGATAG + 9420
TAATCAAGTCGCCCTATCCAGATAACCTTGACTGACATAATAAACCGCACTATC

b L V S A G Y R F Y G T D L Y L Y W R D S -
9421 TATTCTAGCCTATCCATTAACTCTAGTACCGATTGGATAACCAACCATGGTGAACCTAC + 9480
ATAAGATCGGATAAGTAAATTGAGATCATGGCTAACCTAATGGTGGTACCTGGGATG

b I L A Y P F N S S T D W I T N H G D P T -
9481 GACTTGGTATTATAACGAAATAATGGCTCAGGATGGCTATGATGTCCTACTGTGGGT + 9540
CTGAACCATATAATTGCTTTTACCGAGTCTACCGATACTACAGAAGTGAACCCCCA

b T W Y Y N G N M A Q D G Y D V F T V G V -
9541 TGGTGTAAACGGGATCCTGGATGAGCAACGGCTACTAGATTATGGAGAGCAT + 9600
ACCACATTGGCCCTAGGACCATGCCCTACTCTGGTGGCGATGATCTAAATACTCGTA

b G V N G D P G T D E A T A T R F M Q S I -
9601 CTCTAGTTCTCTGACAACCTACACTAACGTAACGAGATCCTCAGATTITACAAGAATT + 9660
GAGATCAAGGGGACTGTTGATGCTAGGTAGAGTCAAGTCTAAATGTCCTAA

Figure 101W

b S S S P D N Y T N V A D P S Q I L Q E L -
 GAATCGCTACTTCTATACTATGCTCAATGAGAAGAAATTCTATGAAAATGGTACCGATTAC
 CTTAGCGATGAAGATAATGATAGCAGTTACTCTTCTATGCTTTACCATGGCTAATG 9720

b N R Y F Y T I V N E K K S I E N G T I T -
 AGACCCGATGGGTGAACTAATTGATTTCCAATTGGGAGCAGATGGAAGGTTTGATGCCAGC
 TCTGGGCTACCCACTTGATTAACCTAAAGTTAACCTCTACCTTCACCTCAAACTAGGTGCG 9780

b D P M G E L I D F Q L G A D G R F D P A -
 GGATTACACTTTAACTGCACAAAGATGGTAGTTGTTGGGAAATAATGTCCTACTGGGGG
 CCTTAATGTTGAAATTGACGTTTGCTACCATCAAACCACTTATTACAGGGATGACCCCC 9840

b D Y T L T A N D G S S L V N N V P T G G -
 ACCACAAATGATGGCTGGCTTGCCTAAAAATGCAAAGGGTTCTATGATAACGACTGAGAA
 TGGTGTCTTACTACACCGAACGATTTTACAGTTTACAAGATACTATGCTGACTCTT 9900

b P Q N D G G L I K N A K V F Y D T T E K -
 AAGGATTCGTTGTAACAGGGTTGGCTACCTTGGAACGGGTGGAAAAAGTTACATTGACTTATAA
 TTCCCTAAGCACATTGTCACAAACATGGAACCTTGCCCACCTTTCAATGTAACCTGAAATT 9960

b R I R V T G L Y L G T G E K V T L T Y N -
 TGTTGCTTGAAATGACCAATTGTAAGCAAATAATTCTATGACACGAAATGGTCGAAACAC
 ACAAGGCAACTTACTGGTTAACATTGCTTAAAGATACTGTTACCGCTTACCGCTTGTTC 10020

b V R L N D Q F V S N K F Y D T N G R T T -
 CCTACACCCCTAAGGAAGTAGAAAAGAACACAGTGGCGGACTTCCGATTCCTAAGATTCCG 10080

Figure 101X

GGATGTGGGATTCCCTCATCTTCTGTCACGGCTGAGGGCTAAGGTTCTAAAGC

L H P K E V E K N T V R D F P I P K I R -

10081 TGATGAGAAAGTATCCAGAAATCACAAATTCCAAAAGAGAAAAACTGGTGAATTGAA
ACTACATGCTTCAAGGTCTTAGGTCTTAGGTAAAGGTTCCTCTTTTGAAACCCTTAAC

D V R K Y P E I T I P K E K K L G E I E -

10141 GTTTATAAGATCAAATAAGAAATGATAAAAACCCTGGAGAGATGGGGTCTTGTGTTCA
CAATYATTCTAGTTATTCTRACTATTTTGGTGTACTCTCTACGCCAGAAATCGAAGT

F I K I N K N D K . K P L R D A V F S L Q -

10201 AAAACAACTCGGATTATCCAGATATTATGGCTATTGATCATAATGGCACTTATCA
TTTGTGTAGGCCCTATAGGTCTATAACCTGATACACTAGTTTACCGTGAATAGT

K Q H P D Y P D I Y G A I D Q N G T Y Q -

10261 AAATGTGAGAACACGGTGAAGATGGTAAGTGTGACCTTAAAATCTGTCAGATGGAAATA
TTTACACTCTGCCCCACTTCACTTCAACTGGAAATTTTAGACAGTCTACCTTTAT

N V R T G E D G K L T F K N L S D G K Y -

10321 TCGATTATTGAAATTGACCAAGCTGGTATAACCCGTTCAAATAAGCTATCGT
ACCTATAACCTTCAAGCTGTGACCTATTTGGCAAGTTTATCGGATAGCA

R L F E N S E P A G Y K P V Q N K P I V -

10381 TGCCTTCAAAATAGTAAATGGAGAAAGTCAGAGATGACTCAATCGTTCACAGATAT
ACGGAGTTTATCAATTACCTTCACTGTTAGCAAGGTGTCTATA

A F Q I V N G E V R D V T S I V P Q D I -

PCT/US05/027239 276/487

Figure 101Y

b

10441	ACCAAGGGTTACGAGTTACGAATGATAAGCACTATATCACAAATTGAGCCAAATTCCCTCC	
	TGGTCGCCAATGCTCAAATGGCTTACTATCGTGTATAATAGTGTACTCGGTAAAGGAGG	+ 10500
b	P A G Y E F T N D K H Y I T N E P I P P -	
10501	AAAAAGAGAATACTCCTCGAACCTGGTGTATCGGAATGGTGCCTATCTGATAGGTG	+ 10560
	TTTTTCTCTTATAGGAGCTTGACCATAGCCTTACAAACGTAAGATACTATCCAAC	
b	K R E Y P R T G G I G M L P F Y L I G C -	
10561	CATGATGATGGGGAGGAGTTCTATTATACACACGGAAACATCCGTAAGGTGTAGCAATGAG	+ 10620
	GTTACTACTACCCTCCTCAAGATAATAATGTTGTGCCTTTGTAGGCCATTACATCGTTACTC	
b	M M G G V L L Y T R K H P *	
10621	AAATGATAATACTCGATACTCTGAGCGATACTTTAAGGAAGTAGCACTCAAGAAGAGATT	+ 10680
	TTTACTATTAGCTATGAGACTCGCTATGAAAATTCTTCATCGTAGTTCTCTCTAA	
10681	AAGTTTACTTGGTGAACAGTTTCTTGCCAAAGTAACCACCATGGAAAGGGGAGATG	+ 10740
	TTCAAATGAAACCCTTTGTCAAAAGAAGGGTTCAATTGGGTAACTTTCCCTCTAC	
10741	TTTTGAAAAACTTGCACAGAAAAGGATATTATTGTCATGTAAATTCAATTACATTGC	+ 10800
	AAAAGCTTTGAACTGTCCTTCTTAATAACAGTACACATTAAGTAATTGTAACG	
10801	TCACAGTTGATTAAAGAGATAATGAAATAAGGAGAAATCATGAAATCAATCAACAAATT	+ 10860
	AGTGTCAACTAAATTCTCTATACTTATTCCCTTTAGTACTTTAGTTAGTTAGTTAGTTAAAA	

c

M K S I N K F L - orf4_670, homologue of sp0463, LPXTRG
TAACATATGCTTGTGCTTATTACTGACAGCGAGTAGCCTGTTAGCTGAAACAGTTT

277/487

Figure 101Z

10861 ATTGTACGAAAGGACCGAATAATGACTGTGGTCATGGACAAAGTGCAGTTGCAAA
C T M L A A L L T A S S L F S A A T V F -
TTCGGGGACAATGTTAGTACAGCACCAATGCTGTACTAAAACCTTAACATCCATA
10921 AACGCCGCGCTGTTACAATCATGTCGGCTACGACAATGATTGAAATTGTAGGTAT
C A A D N V S T A P D A V T K T L T I H K -
AGTTTACUTGCTTCAGAACATGGATTAAAAGACUTGGGATACAAACGGTCCCTAAACGGATATG
10981 TCAATGACGGAGAGCTCTACTAAATTCTGAACCCTATGTTGCAGGATTCTCTATAC
C L L S E D D I K T W D T N G P K G Y D -
ATGGAACCTCAATCTAGTTAAAGATTAAACTGGAGTTGACTGAGGTGAGCTGAGGAATTCCAATG
11041 TACCTTGAGTTAGATCAAATTCTAAATTGACCTCAACATCGACTCCCTTAAGGTTAC
C G T Q S S L K D L T G V V A E E I P N V -
TATACTTTGAAATTACAAAAAGTATAATTGACTGATGGTAAGAAAAAAATCTTAAG
11101 ATATGAAACTTAATGTTTCATATTAACCTGACTACCAATTCCCTTTCTTAGAATTTC
C Y F E L Q K Y N L T D G K E K E N L K D -
ATGATAGTAAATGGACAAACAGTTCACTGGGGTTGACAACTAAAGATGGACTTAATG
11161 TACTATCAATTACCTGTTGTCAGTACCCAAACTGTTGATTTCACCTGAAATTAAAC
C D S K W T T V H G G L T T K D G L K I E -
AAACCAAGTACTCTTAAGGTGTGTATCGTATTGCTGAGGATAGAACAAAGACTACCTATG
11221 TTGGGTCAATGAGAATTCCACACATAGCATAAAGCACTCCATCTGTTCTGAGGGATAC
C T S T L K G V Y R I R E D R T K T Y V -

PCT/US605/27234

278/487

Figure 101AA

TTTGGTCCTAAATGGGCAAGTATTAAACGGTTCAAAGGCCGTACTCTGCTCTTGTAACTCTTC
 11281 ----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11340
 AACCGGATTACCGTTCATATTGTCCAAGTTTCGGCATGGGACGAGAACATTGAGAAG

 C G P N G Q V L T G S K A V P A L V T L P -

 CACTTGTGTTAACATAATGGTACAGTAATTGATGCACATGTTTCCCTAAATTCAATA
 11341 ----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11400
 GTGAACAATTGTTATTACCATGTCATTAACTACGGTGTACAAAAGGGATTTTAAAGTATAT

 C L V N N N G T V I D A H V F P K N S Y N -

 ATAAACCAGTGTGATATAAGAAATTGGCTGATACTTTGAATTATAACGGATCATAATGGTC
 11401 ----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11460
 TATTGGTCAACATCTATTTCCTAACCGACTATGAAACTTAAATTGGCTGATTTCACAG

 C K P V V D K R I A D T L N Y N D Q N G L -

 TGTCTATCGGTACTAAATCCCATAATGGTTAAATAACAACAAATTCCAAGTAAATGAAACAT
 11461 ----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11520
 ACAGATAGCCATGATTAGGGTATACAAACAATTATGGTTGTTAAAGGTTCATTAACGTTGTA

 C S I G T K I P Y V N T T I P S N A T F -

 TTGCAACTTICATTGGTCAGATGAAATGACAGAAGGTCAACTTATAATGGAGATGTA
 11521 ----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11580
 AACGTTGAAAGTAAACCGGCTACTTTACTGTTCCAGATTGAAATAACTCTACATT

 C A T S F W S D E M T E G L T Y N E D V T -

 CATTACTTGTGAAATTGTGCTATGGATCAAGTGATTATGAAAGTCACTAAGGAATA
 11581 ----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11640
 GTTAATGAAACTTAAATTACATCGATAACCTGAGTTCGACTAATACTTCAGTGATTTCCTTTAT

 C I T L N N V A M D Q A D Y E V T K G N N -

 ATGGCTTAACTTAAACAGAAGCAGGGTTAGCTAAATTAAATGCTTAAAGGTGAGATGCGAG
 11641 ----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11700
 TACCGAAATTGAAATTAAATTGCTTCAAGTCGTTCAAAATCGATTTTAAATTACCATTCCTACGTC

PCT/US05/27239 279/487

Figure 101AB

C G F N L K L T E A G L A K I N G K D A D -
 ACCAAAAATCCAAATTACTTACTCAGCTTACTTTGAACTCACTTGGTGTGAGACATT
 11701 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11760

C Q K I Q I T Y S A T L N S L A V A D I P -
 CTGAAAGTAACGATAATTACATTAATCGGAATCATCAAGATCATGGAAATACTCCAA
 11761 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11820

C E S N D I T Y H Y G N H Q D H G N T P K -
 AACCAACTAAACCDAAATAATGGTCAAATTACAGTAACDAAAGACATGGACAGTCAACCTG
 11821 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11880

C P T K P N N G Q I T V T K T W D S Q P A -
 CTCCTGGGGTAAAGGGACTTCAACTTGTAAATGCCAAGACTGGTGAGAAAGTCG
 11881 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 11940

C P E G V K A T V Q L V N A K T G E K V G -
 GTGCCTCCCTGTAGAACTTCAGAAATAATGGACATATACTTGGAGTGTAGATAATT
 11941 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 12000

C A P V E L S E N N W T Y T W S G L D N S -
 CTATTGAAATACAAAGTTGAAAGAAATATAATGGATAACTCAGCTGAAATACAGTAGAGA
 12001 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 12060

C T E V K V E E E V N G Y S A E V T V E S -
 GCAGGGAAAGTGGGGTAAANAAACTGGAAAGATAATACCAGCTCCAAATCCTG
 12061 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 12120

PCT/LIS05/27239

280/487

Figure 101AC

CGTTTCCTTCAACCCCCATTTCGACCTTCTATTGGGTCAAGGGTTAGTTAGGAC
 C K G K L G V K N W K D N N P A P I N P E -
 AAGAACACGTTAAACATACGGTAAAAGTTGTCAAAGTAGACCCAAAAGATACTC
 12121 TTCTTGTTGCACATTTGTATGCCATTTCACAGTTCATCTGGTTTCATCTATGAG
 C E P R V K T Y G K K F V K V D Q K D T R -
 GTCTAGAAAATGCCAGTTGTTGTTAAAAAGCAGATAGCAATAAATATTGCCCTTA
 12181 CAGATCTTTACGGTCAAGCAACAATTTCGTCATCGTTATTATAACGGAAT
 C L E N A Q F V V K A D S N K Y I A F K -
 AGTCAACTGACACAAAGCTGCAGATGAAAAAGCAGCACAACGTGCAAAACAAAAATTGG
 12241 TCAGTTGACGTGTTGTTGACGGCTACTTTCTGCTGCGTTGACGTTTGTTTAACC
 C S T A Q Q A A D E K A A T A K Q K L D -
 ATGCAGGGTAGCAGCTTACACAAATGCTGCAGATAAGCAAGGCCGCTCAAGCTCTAGTAG
 12301 TACGTGCCATCGTCAATTGTTACGACGTCTATTGTTGAGTCGAGATCATC
 C A A V A A Y T N A A D K Q A A Q A L V D -
 ATCAAGCACAGCAAGAATAACAAATGCTACAAAGAAGCCAAATTGGTTATGGTTGAAG
 12361 TAGTTGTTGTTACATCGAAATTGTTCACTGGTTAAACCAAAATACAACTTC
 C Q A Q Q E Y N V A Y K E A K F G Y V E V -
 TAGCTGGAAAAGATGAAGCAATGGTTCTTACTCTTAATACGGATGGTCAAATTCCAATTT
 12421 ATCGACCTTTCTACTTCGTTACCAAGAATGAAGATTGCTTACCAAGTTAAGGTTAAA
 C A G K D E A M V L T S N T D G Q F Q I S -

PCT/US05/2723281/487

Figure 101A

PCT/US05/27239 82/487

Figure 101AE

a V M A L C F S L V W G H A V Q A Q E D -
 CACACGGTTGGTCCTTGCATTGGAGAACTATCAGGAGGTCTCCACCAATTCAGTTAACGGTAAACGGCA
 129001 GTGTGCAACCAGAACGGTTAACCTCTTGATAATGTCATGCCATTACGGTTAACGGTAAACGGTAAACGGCA 12960

a H T L V L Q L E N Y Q E V V S Q L P S R -
 GATGGTCATCGGTGGCAAGTATGGATGGATGATTCTATGATGATGGGTG
 12961 CTACCCAGTAGGCCAACGGTTCAACCTACTAAGCATAAAGGATAACTACAGCCCCAC 13020

a D G H R L Q V W K L D D S Y S Y D D R V -
 CAAATTGTAAGAGACATTCGATTGGATGGATGAGAATAAACTTTCTTCATAAAAGACT
 13021 GTTTAACATTCTCTGAACGTAAGCACCCTACTCTTATTGAAAGAAAGTTTCTGA 13080

a Q I V R D L H S W D E N K L S S F K K T -
 TCGTTTGAGATGACCTTCCTTGAGAATCAGGATTGAAAGTATCCTCATATTCCAAATGGCTTT
 13081 AGCAAACTCTAATCTGAAGGAACCTCTTCTAGTCATAACTCTATAGAGTTAAAGGTAAACAGAA 13140

a S F E M T F L E N Q I E V S H I P N G L -
 TACTATGTTGGCTCTTATTATCCAGACGGATGGGATGGGTTTCTTATCCAGGTGAATTTCCTTT
 13141 ATGATACAAAGCGAGATAATAGGTCTGCCTAACGCCAACGAAATAGGTCAACTTAAGAAAAA 13200

a Y Y V R S I I Q T D A V S Y P A E F L F -
 GAAATGACAGATCAAACGGTAGAGGCCCTTGGTCATTGAGCGAAAAAAACAGATACAATG
 13201 CTTTACTGGCTAGTTGCCATCTGGAAACACAGTAACATCGCTTTTTGGCTATGGTAC 13260

a E M T D Q T V E P L V I V A K K T D T M -
 ACAACAAAGGGTAGGGCTGATAAGGTGGATCAAGGCAACATCGCTTGGGGTGCAGGC
 13261 13320

PCT/US05/27239 283/487

Figure 101AF

a T T K V K L I K V D Q D H N R L E G V G -
 TTTAAATTGGTATTCAGTAGGAGATGGTCTGAAAAGAGGTTCCTGATTGGAGA
 13381 +-----+-----+-----+-----+-----+-----+-----+-----+ 13380
 AAATTAAACCATAGTCATCGTTCTACCAZAGACTTTCTCCAAAGGAACTAACCTCTT

a F K L V S V A R D G S E K E V P L I G E -
 TACCGTTACAGTTCTTCTGGTCAAGTAGGAGAACTCTATACTGATAAAAATGGAG
 13381 +-----+-----+-----+-----+-----+-----+-----+-----+ 13440
 ATGGCAATATGTCAAAGAACGAGTTCAATCCCCTCTTGAGAGATATGACTTTTACCTCTC

a Y R Y S S S G Q V G R T L Y T D K N G E -
 ATTTTTGTGACAAMCTTCTCTTGGAAACTATCGTTCAAGGAGGTGGGCCACTGGCA
 13441 +-----+-----+-----+-----+-----+-----+-----+-----+ 13500
 TAAAAACACTGTGTTAGAAGGAAACCCTTGATAGCAAAGTTCTCCACCTCGGTGACCGT

a I F V T N L P L G N Y R F K E V E P L A -
 GGCTATGCTGTACGAGCTGGATAACGGATGTCAGCTGGTAGATCATCGCTGGTGACG
 13501 +-----+-----+-----+-----+-----+-----+-----+-----+ 13560
 CCGATAAGACAATGCTGGCAGCCTATGCTCAAGGTGACCATCTAGTAGTCGACCACTGC

a G Y A V T P L D T D V Q L V D H Q L V T -
 ATTACGGTTGTCATTAGAAATTACCAACGTGGCAATGTCAGTTTATGAAAGGTGGATGGT
 13561 +-----+-----+-----+-----+-----+-----+-----+-----+ 13620
 TAATGCGAACAGTTAGTCCTTATGGCACCGTTACAACCTGAAATACTTCCACACTACCA

a I T V V N Q K L P R G N V D F M K V D G -
 CGGACCAATACCTCTCTTCAGGGCAATGTCAAAGTCATGAAAGAAGAAAGGGACAC
 13621 +-----+-----+-----+-----+-----+-----+-----+-----+ 13680
 GCCTGGTTATGGAGGAAGTTCCCGTTACAAAGTTCAGTACTTTCGCGCTGTG

a R T N T S L Q G A M F K V M K E S G H -

Figure 101AG

13681	a	Y T P V L Q N G K E V V V T S G K D G R -	TATACTCCTGTTCTTCAAATGGTAAGGAAGTAGTTAACATCAGGAAAGATGGTCGT ATATGAGGACAAAGTTTACCATTCCTTCATCAACATTGTAGTCCTTCTACCGAGCA	13740
13741	a	F R V E G L E Y G T Y Y L W E L Q A P T -	TTCGGAGTGGAGGTCTAGAGTATGGGACATACTATTATGGAGCTCCAAGGTCACA AAGGCTCACCTTCCAGATCTCATACCTGTATGATAAAATACCCTGAGGTTCGAGGTGA	13800
13801	a	G Y V L T S P V S F T I G K D T R K E -	GGTATGGTCAATTAAACATGCCCTGTTCCCTTACAATGGAAAGATACTCGTAAGGAA CCATAACAAGTTAAATGTAGCGGACAAGGAATGTAGCCCTTCTATGAGCATTCCTT	13860
13861	a	L V T V V K N N K R P R I D V P D T G E -	CGGGTAACAGTGGTTAAAATAACAAGGCACCGATTGATTGCCAGATACAGGGAA GACCATGGTCACCAATTTTTATTGGTGGCTTAACCTACAGGTCTATGTCCTT	13920
13921	a	E T L Y I L M L V A I L L F G S G Y Y L -	GAACCTTGATATCTGATGGCTTGTGCCATTGTTGGTAGTTGGTTATTATCTT CTTGGAACATATAGAACCTACGAACACCGTTAACACAACCATCACCATTATAGAA	13980
13981	a	T K K P N N * *	ACGAAAAACCAATAACTGATATTCAATGTACATCATTATGAAAAGATAAGCAGGCTGA TGCTTTGGTTTATTGGACTATAAGTTACATGTAGTAATACTTTCTATCGTCGCCGACT	14040
14041	a	T K K P N N * *	AGGGAAAGACCAAGAGTACTCTGAGGTGATGTTAACATGGAATCATGGTATGGCATGAA TCCTTCTGGCTCATGAGACTCCACTAACATTAGTCCTTACCAATTAGTCCTTACCCGTACTT	14100

Figure 101AH

14101	TCACAAATAACGGATAATGAGGCTGGCAGATTGTGCCAGCTCATTTGGTATTGTTG	14160
	AGTGTATTGGCTATACTCCGACCCGTCTAACACGGTGGAGTAACACCCAATAAAC	
14161	TAACACGATAGGACTGGTCTGGTAATCATTAGGATGGACAGGACTGGATTCTGATT	14220
	ATTTTGCTATCCGACCCATTAGTAAATCCTTACCTGTCTGACCCCTAAGCTAA	
14221	AAAAATGGATGGAAATCAGAAAATGAGATTCTCGTTCTCTAGCAGATAGGAT	14280
	ATTTTACCTTACCACTTAGTCCTTTACTCTAAAGGAGAACGAAATCGCTTATCCTA	
14281	TGTCCTGTTAGGAAAGCGATAAAATGATGAGTTGAAGATAAAGGGATGCTGATAAAAT	14340
	ACAGACAATCCCTTTCGCTTAACTACTCAAACCTCTATTTCCTACGGACTTTTTA	
b	M L I K M -orf6_670, homologue of sp0466, sortase	-
14341	GGTAAAAACAAAAAGCAAAACGAAATAATCCTTATTAGGGGGTATTTCATTTGG	14400
	CCATTTTGTCTTTCGTTTGTCTTATTAGGGATAATCCTCACCCATAAAAGTAACC	
14401	AATGGGGTAATGGGTATCGCGTCTCGCTTGTATTATCGAGTGGATGGAATCAATCA	14460
	TTACCGCCATTACGGCATGGGACCAACAGGGAAACATAATAGCTCACCTTGTATTAGT	
b	V K T K Q R N N L L G V V F F I G -	-
14461	ACAAAATTGCTGACTTTGATAAGGAAAAAGCAAGTGGATGAGGCTGACATTGATGAAACG	14520
	TGTTTAACGACTGAAACTATTCTTTCTGTTGAAATAATGAGCTAACACTACTTGC	
b	Q I A D F D K E K A T L D E A D I D E R -	-
	AATGAAATTGGACAAGCCCTCAATGACCTCTTGAATAATGAGCTGAGTGGCAGTCCTTG	

PCT/US05/27239

286/487

Figure 101AI

b M K L A Q A F N D S L N N V V S G D P W - + 14580
 14581 TTA CTT TAA CCG GTT CGG AA GT TACT GT GAA AAA ACT TTACATCACTCACCGTAGGAAC
 GT CGG AAG AAA ATG AAG AAA AGGGCGAGCAGAGTATGCACGTATGTTAGAAATCCATGA
 CAG CCT CTT TACT TCT TTT TCC CG CT CG TCT CAT AC GT GC AT CA AA AT CT T TAG GT TACT + 14640

b S E M K K G R A E Y A R M L E I H E - + 14641
 GCG GAT GGG C AT GT GGA AA AT CCC CG TT AT T GAC GT GG AT T T GCG GGT T AT GCG GGT AC
 CGC CT ACC CG TA ACC TT TAG GGC A AT A ACT GC ACCT AA CG G C C A AA AT ACC G ACC AT G + 14700

b R M G H V E I P V I D V D L P V Y A G T - + 14701
 TG CT GA AAG GT TATT GAG CAAGGGCTGGGCATCTAGGGAAACTTCTCTGCCGATCGG
 ACC ACT TCT CC CATA AAC GT CG TCCC GAC CG GT AG AT C TCC CT TGA AAG AC GG CT AG CC + 14760

b A E E V L Q Q G A G H L E G T S L P I G - + 14761
 AG GCA ATT CG ACC AT GCG GT GATT AC GGC AC AT A CAG GT T T G C C A A CAG CT AA G AT GT T
 TCG GT TA AG CT GG GT AC G C C ACT A AT G C C GT GT AT G T C C A A A C G G T T G C G AT T C T A C A A + 14820

b G N S T H A V I T A H T G L P T A K M F - + 14821
 TAC GG AT T GACC AA ACT TA AAG GT GGG ATA AAG T T T AT G T G C A C A A T A T C A A G G A A G T
 AT G C C T A A A C T G G T T G A A T T C A A C C C T A T T C A A A T A C A C G T G T T A T A G T T C C T T C A + 14880

b T D L T K L K V G D K F Y V H N I K E V - + 14881
 GAT GGC CT AT C AAG GT GG AT C AAG TA AAG GT GATT GAG CC GAC G A ACT T T G A T G A T T T A T T
 CTAC CG G AT A G T T CAC CT AG T C AT T C C A C T A A C T C G G C T G C T G A A A C T A A T A A A + 14940

b M A Y Q V D Q V K V I E P T N F D D L L -

PCT/US105/22239

287/487

Figure 101AU

PCT/US05/27239

288/487

Figure 101AK

C S N E V I K E F D E T V S Q M D K A E L -
 AGTCAAACGAGGTATTAAAGAGTTTGATGAGACGGTTCCAGATGGATAAGCGAAC
 15361 TCAGTTGCTCCAATAATTCTCAAACACTCTGCCTAACCTATTCCGTCTTG
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 15420

C E E R W R L A Q A F N A T L K P S E I L -
 TTGAGGAGCCTTGGCCTTGGCTCAAGCCTTCAATGGACCTTGAACCATTGAAATTC
 15421 AACTCTCGCAACCGGAACCGAGTTGGAAAGTTACGGCTGGAAACTTGGTAGACTTTAAG
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 15480

C D P F T E Q E K K G V S E Y A N M L K -
 AGGTCCATGAGGGATTGGCTATGTGGAAATTCTGCATTGATCAGGAATTCGGATGT
 15481 AACTAGGAAATGTCTCGTTCTCTTCTTCTTCCGGAGGTCTTACGGTTATACGGATT
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 15540

C V H E R I G Y V E I P A I D Q E I P M Y -
 ATGTCGGAACGAGGTGAGGAATTCTTCAGGAGGGCGAGGATTGCTAGAGGGGCTTCGT
 15541 TACAGCCTTGCCTACTCCTTAAGAAGTCTCCGGCTCCTAACGATCTCCCTCGAAGCA
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 15600

C V G T S E E I L Q K G A G L L E G A S L -
 TACCGGTTGGTGGAAATAACCCACACAGTTGTCACTGCTCATAGGGATTACCGACGG
 15601 ATGGCCAACCAACTTTATGGCTTACAGTGAGGTATCTCCCTAATGGCTGCC
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 15660

C P V G G E N T H T V V T A H R G L P T A -
 CAGAAACTGTTAGTCAAATTGGATAAGATGAAAAAGGGGATGTCTTATCTTCAGGTT
 15661 GTCTTGACAATCAGTTAACCTTACTTCTACTTCTACAGAAATAGAAGTGC
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 15780

C P V G G E N T H T V V T A H R G L P T A -
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 15720

C P V G G E N T H T V V T A H R G L P T A -
 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 15721

PCT/US05/27239

289/487

Figure 10A

c E L F S Q L D K M K G D V F Y L H V L -
 TAGACCAGGTGTTGGCCTACCAACTGATCAGATTGACGGTGTAGCCAAATGGACTTGC
 15781 ATCTGTCACCAACCGGAACTTCACCTAGCTAAAACCTGCAACTGGTTTACTGAAAC

c D Q V L A Y Q V D Q I L T V E P N D F E -
 AGCCTGTTGATTCAACATGGGAGATTATGCCACCTTGTGACTCTGTACACCGTATA
 15841 TCGGACAGAACTAACATTGTACCCCTCTATACTGCTGGAAACAACCTGACATGTGGCATAT

c P V L I Q H G E D Y A T L L T C T P Y M -
 TGATTAAACAGTCATCTGTCTGTGTGTACGTGGGAAGGGATTACGGCACCAAATTG
 15901 ACTAAATTGTCAGTAGCACACAACTGACCCATGGCCATATGCCATATGCCGTGGTTAAC

c I N S H R L L V R G K R I P Y T A P I A -
 CAGAGCCAAATCGGGGTCAAGAGAGCTGGCAATTCTGGTTGTGTGTATTGCTAGGG
 15961 GTCTCGCTTAACTCGCCACTCTCGCACCGTTAACACCAACAAATAACGATCGCC

c E R N R A V R E R G Q F W L W L L A A -
 CGTTGGTTATGATTCTGGTATTGAGTTACGGGGTGTATCGTCATCGCATTTGTCAAAG
 16021 GCAACCAAATCAACTAAAGACCAATCTGCACTAGCTGGTAAACAGTTTC

c L V M I L V L S Y G V Y R H R R I V K G -
 GGCTTAAGAAACAAATTGGAGGAGCATCTGCTAAAGGCTAAGGTACAGAAATTACTAGGG
 16081 CCCATCTGGTTGGTAACTCCCTCTGAGTTCCGATTTGAGATCTTAAATGATCCC

a M S K A K L Q K L L G -orf8_670, homologue of sp0468, sortase

c L E K Q L E H H V K G *

PCT/US05/27239

290/487

Figure 101AM

Figure 101AM

<pre> TATTTGGCTAATGCTGGTAGCAATTGGGAAATTCCCGTTAATGTTGGGAGATGGGTAA ATTAACGACTAACCCATCGTAACCACTAAGGACAATAACAAAACCCGTCTACACAAAT </pre>	16141
<pre> CAGTCTTCTAGGACAAGTAAAGGTCAATGAGATTTCTAGAAATCTGTGACGGCCGACAGT GTCAGAAATCCTGTTCAATTTCAGTACTCTATAAAGCTTCTAGACACTGCCGGCTGTCA </pre>	16201
<pre> a Q S L G Q V K G H E I F S E S V T A D S - TACCOAGGAAATTGCAACGGTGGCTTGATCAATCAACGGCTGGATTGCAAATGCT </pre>	16261
<pre> ATGGTTCTCGTTAACGGTGGCAAGGAACTAATGTTAGTTAGCTTGGAACTTAAGGTTTTAGCA </pre>	
<pre> a Y Q E Q L Q R S L D Y N Q R L D S Q N R - ATTGTAGATCCTTTTGGGGAAAGGTATGAGGTAAATTACCAAGTGTCTGACGATCCT </pre>	16321
<pre> TAACATCTAGGAAAAACGGCTTCCCATCTCTCCATTAACTGTTACAGACTGCTAGGA </pre>	
<pre> a I V D P F L A E G Y E V N Y Q V S D D P - GATGGCAGTCAACGGCTATTGTGGATTCCGAGTTGGAAATCATGGAGCCAGTTATCTA </pre>	16381
<pre> CTACGTCAGATGCCGATAAACGCTAACGGCTAAACCTTATGACTCTCGGTCAAATAGAT </pre>	
<pre> a D A V Y G Y L S I P S L E I M E P V Y L - GGAGGGGATTACCATCATTAGCAATGGGGTTGGCCCATGTGGATGGGACGCCTCTTCTC </pre>	16441
<pre> CCTCGCTTAATGGTAGTAAATGTTAACCCCAACGGGTACACCTACCCCTGGGAGAAGGA </pre>	
<pre> a G A D Y H H L A M G L A H V D G T P L P - GTTGAGGGAAAGGATTCTGTTCACTGTTAGTGTGGGACCCGTGAGAACCAAGCCTGTC </pre>	16501
<pre> CAACTCCCTTTCCCTAACGAAAGTCACTAACGACCCGTGGCACETCTGGTACAG </pre>	

Figure 101AM

a V E G K G I R S V I A G H R A E P S H V -

TTTTTCGCCATTGGATCAGCTAAAGTGGAGATGGCTCTTTATTATGATAATGGCCAG
16561 AAAAGGGGGTAAACCTAGTCGATTTCACCGAGAATAACTATTACCGGTC
16620

a F F R H L D Q L K V G D A L Y Y D N G Q -

GAAATTGTAGAATATCAGATGATGGACACAGAGATTATTTACCGTGGAAATGGAAAAA
16621 CTTAACATCTTATAGTCTACTACCTGTTCTAAATAATGGCAGGCCCTAACCTTTT
16680

a E I V E Y Q M M D T E I I L P S E W E K -

TITAGAAATCGGTTAGCTCTAAATAATCATGACCTTGATAACCTGCGATCCGATTCCTTAC
16681 AATCTTAGCCAATCGAGATTATAGTACTGGAACTATTGGACGCTAGGCTAAGGATGG
16740

a L E S V S S K N I M T L I T C D P I P T -

TTTAATAAACGCTTATTAGTGAATTGGAAAGCAGTCGCTGTTTATCAAAAATCAGATCCA
16741 AAATTATTTGCGAATAATCACTAAACTTGCCTCAGGACAATAATGTTTAGCTAGGT
16800

a F N K R L L V N F E R V A V Y Q K S D P -

CAAACAGCTGCAGTTGCAGGGTTGCTTACGAAAGAAGGACAATCTGTATCGCGTGT
16801 GTTTGTGACGTCAAGCTCCAAACGAAAATGCTTCTCTGTAGACATAGGCCACAA
16860

a Q T A A V A R V A F T K E G Q S V S R V -

GCAACCTCTCAATGGTTGTACCGTGGCTAGTGGTACTGGCATTTCTGGGAATCCTCTGGTT
16861 CGTTGGAGAGTTACCAACATGGCACCGATCACCAGGCCATAAGACCCCTAGGACAAA
16920

a A T S Q W L Y R G L V V I A F L G I L F -

GTTTTGGAAAGCTAGCACGGTTACTACGAGGGAAATAAAAATGAAAGGAAAGCTA
16921

PCT/US05/22239 292/487

Figure 101AN

CAAAACACCTTCGATCAGCCAAATGATCCTCCCTTATTTCCTTTACTTTCTTCGAT

a V L W K L A R I L R G K *

16981 AGGGTGTTCCTTTCGGGCTCTTCTGAACTCTAGTGGTTGAAAAAAGCTAAGCTCG
 TCCGACAGGAAAAGGCCAGAACAGTTGACATCACCAACTTTTGATTCGAGC

17040 AGAAAGGACAATAATTGTCCTTCCTTGTATTCAGAGCGATAAAATCCGGTTTT
 TCTTTCCTGTTAAACAGGAAGAAAAACTATAAGTCTCGGTATTTTGGCAAAAAA

17101 GAAGTTTCAA
 17112 CTTCAAAAGTT

PCT/US05/27239

293/487

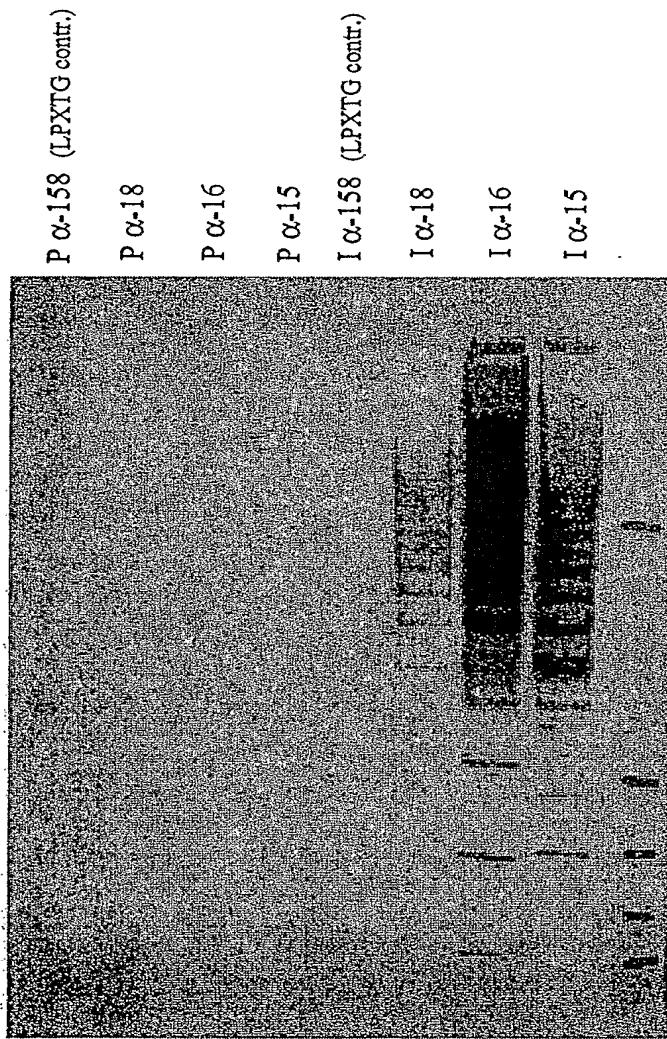
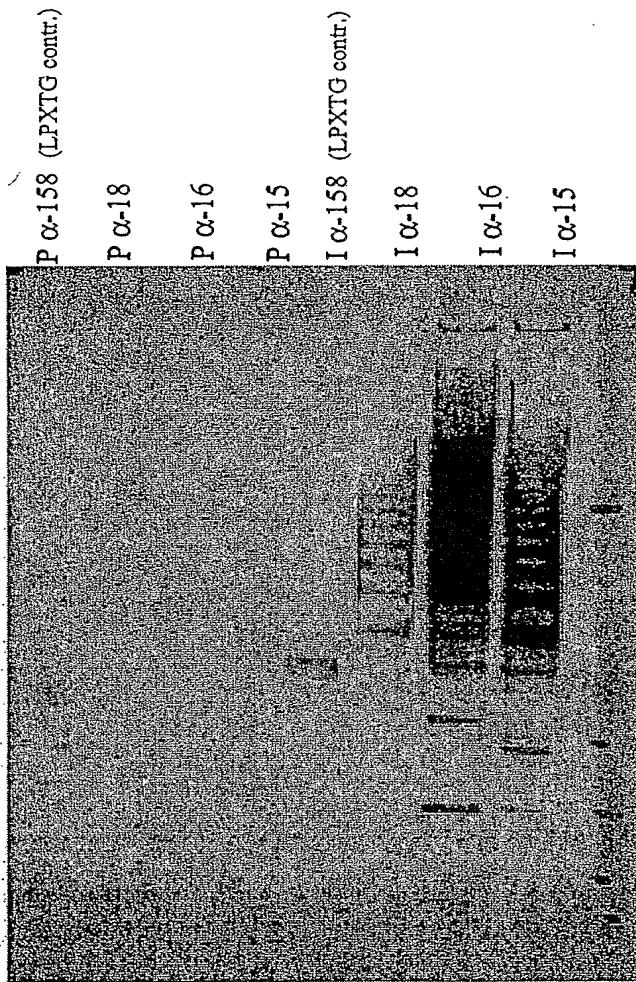
M1, strain 2580

Figure 102

LEGEND:

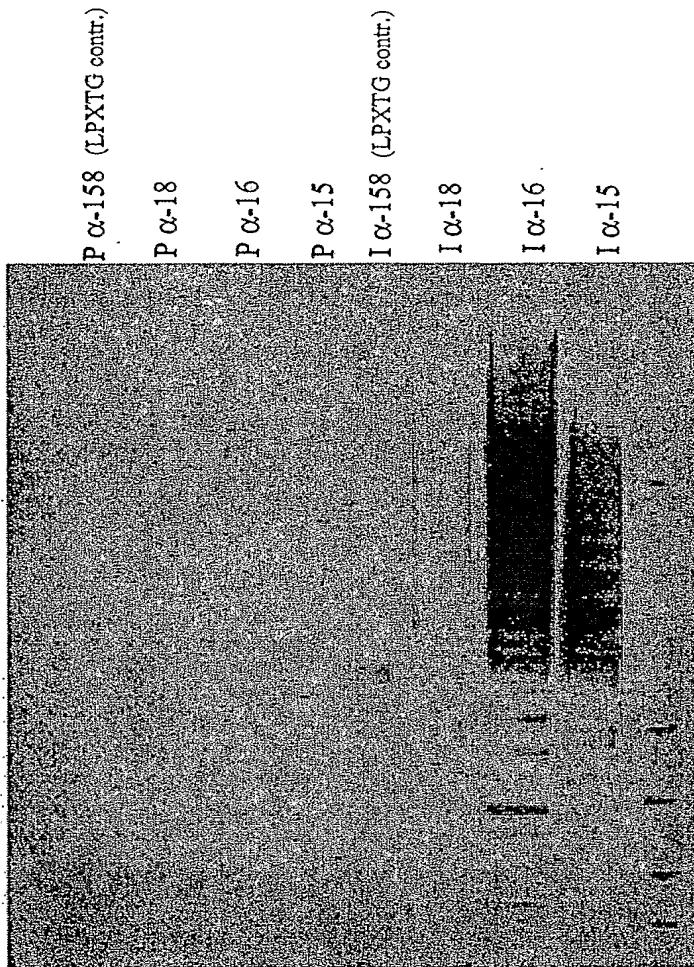
I α -#: immune serum anti-#
P α -#: pre-immune serum anti-#

PCT/US05/27239 294/487

M1, strain 2913**Figure 103****LEGEND:**

I α -#: immune serum anti-#
P α -#: pre-immune serum anti-#

M1, strain 3280

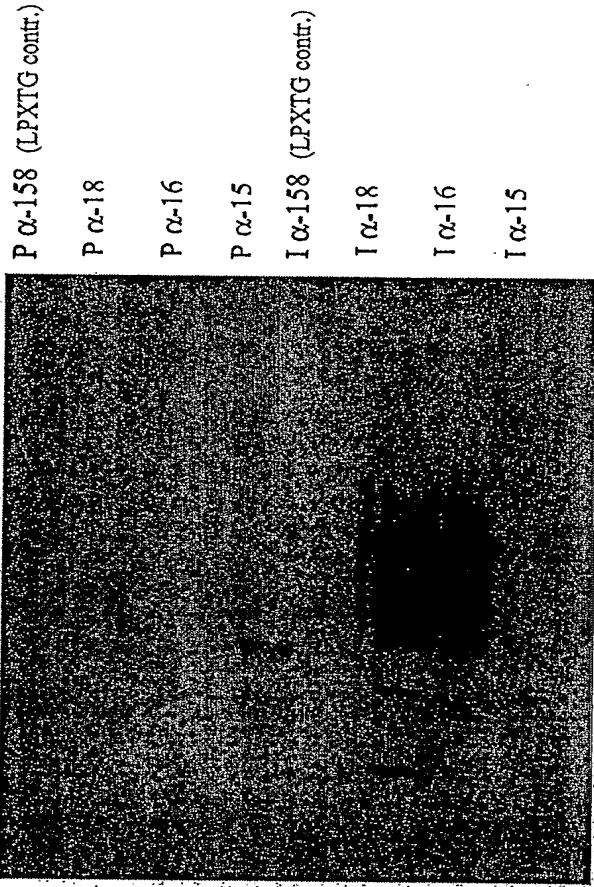


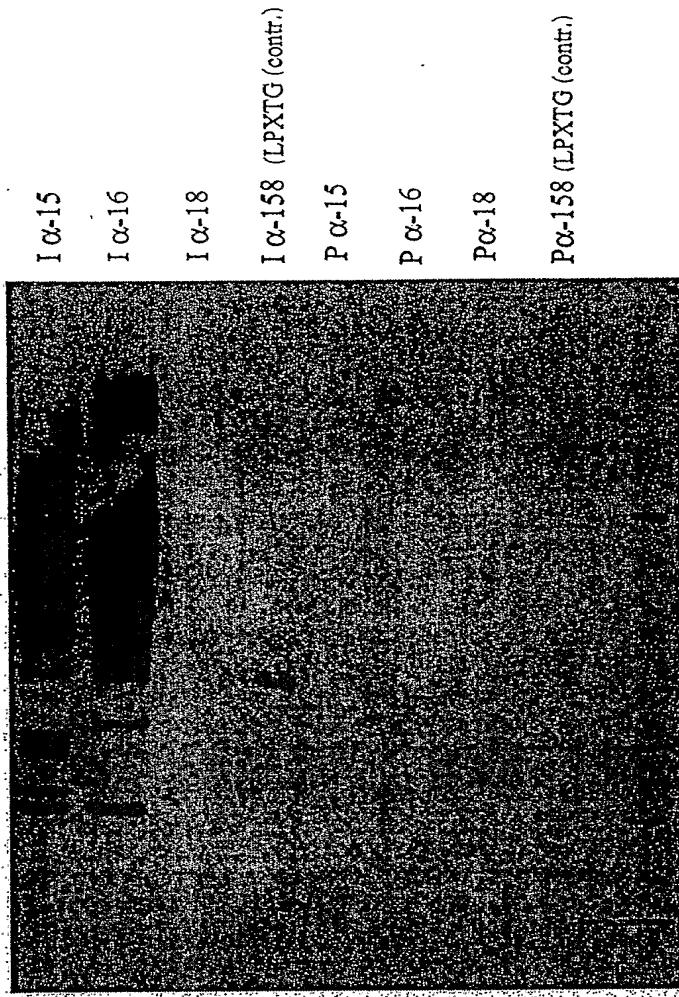
LEGEND:

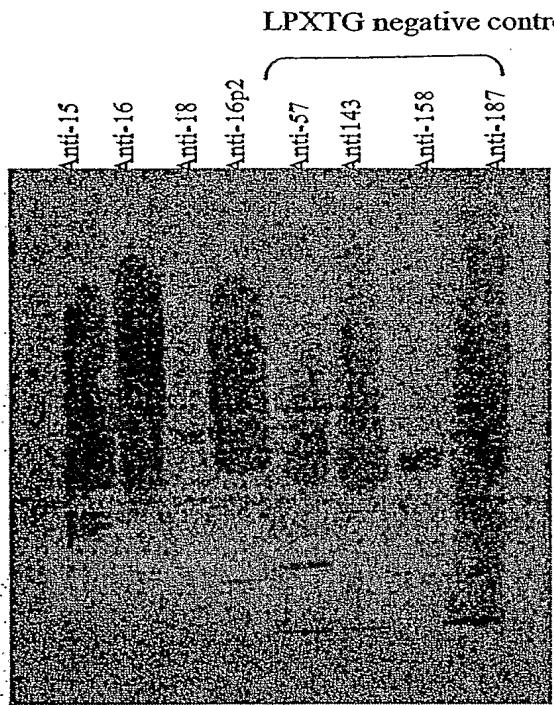
I α -#: immune serum anti-#
P α -#: pre-immune serum anti-#

Figure 104

296/487

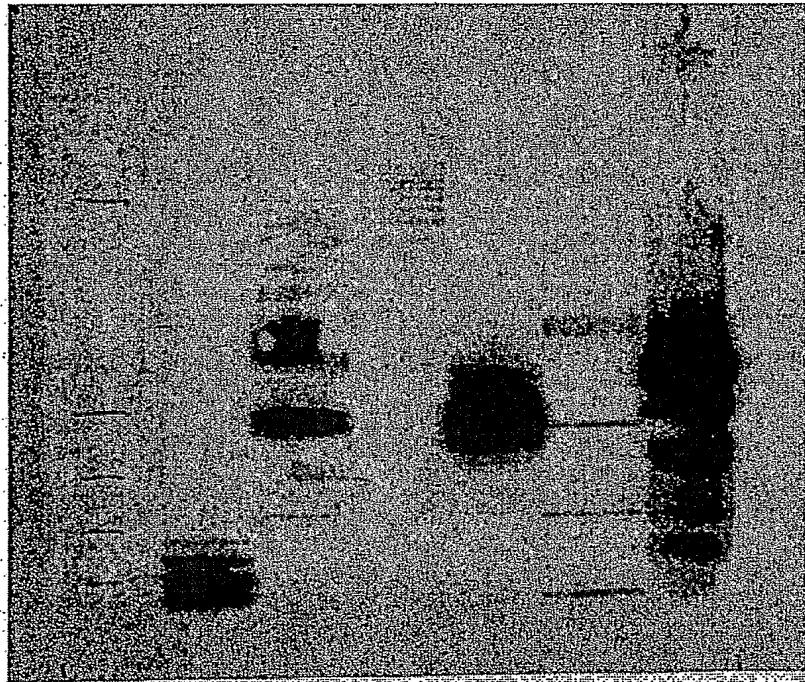
M1 strain 3348**Figure 105**

M1 strain 2719**Figure 106**

Western blot Western blot on fraction enriched in surface proteins of M1 (SF370)**Figure 107**

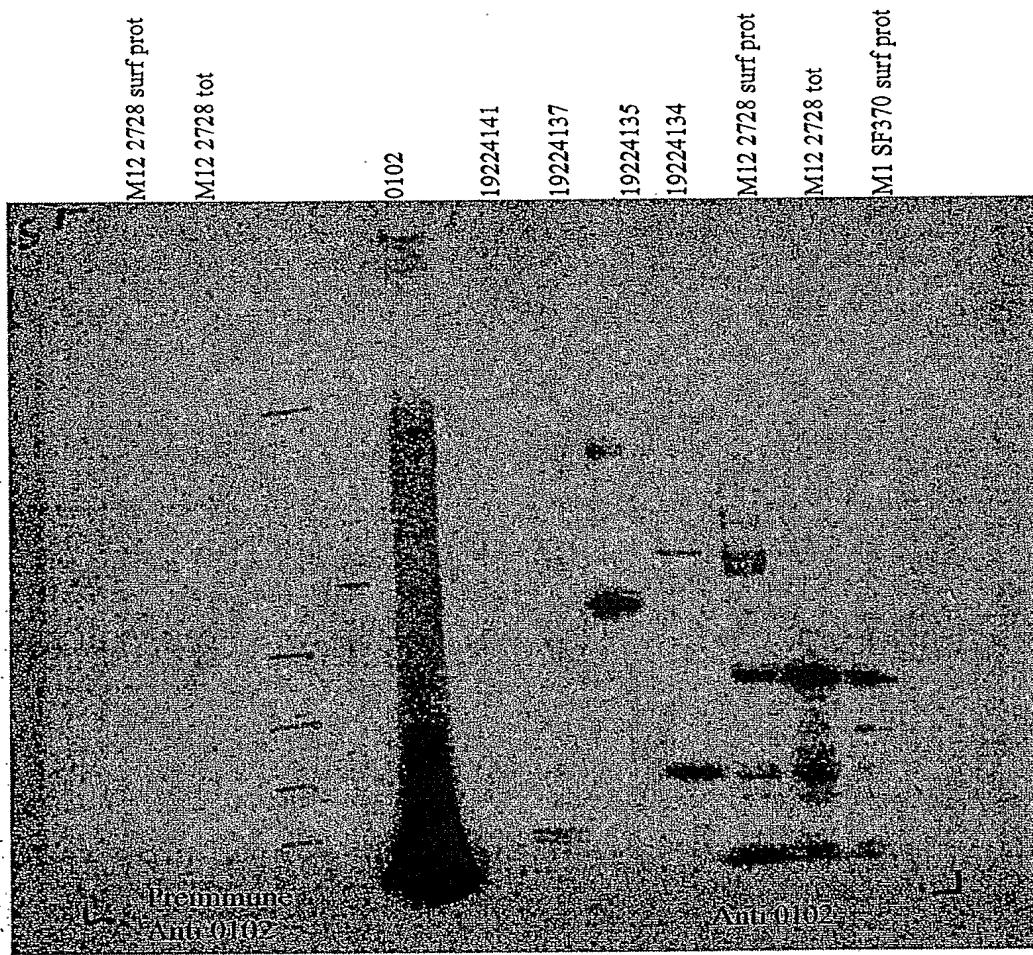
Western blot on fraction enriched in surface proteins of M12 (2728)

Anti-19224134
Anti-19224135
Anti-19224137
Anti-19224141
Anti-0102
Anti-158 (control)



Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2a)"-, the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in M12 (strain isolate 2728). This means that in M12 three out of the five LPKTG proteins form high-molecular weight structures:

Figure 108



Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2bis)"-, the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in protein extracts from M12 (strain isolate 2728.)

Figure 109

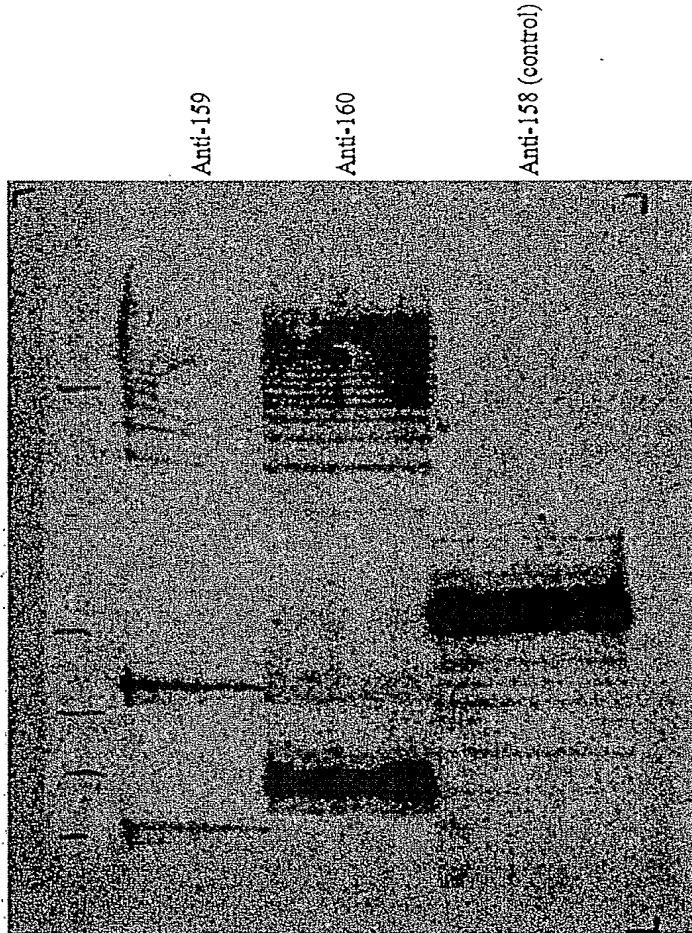
WO 2006/078318

301/487

PCT/US2005/027239

PCT/US05/27239

Western blot on fraction enriched in surface proteins of M6 (2724)



M6 strain isolate 2724

Figure 110

WO 2006/078318

302/487

PCT/US2005/027239

PCT/US05/27239

0.160

0.159

M6 3650 surf prot

M1 SF370 surf prot

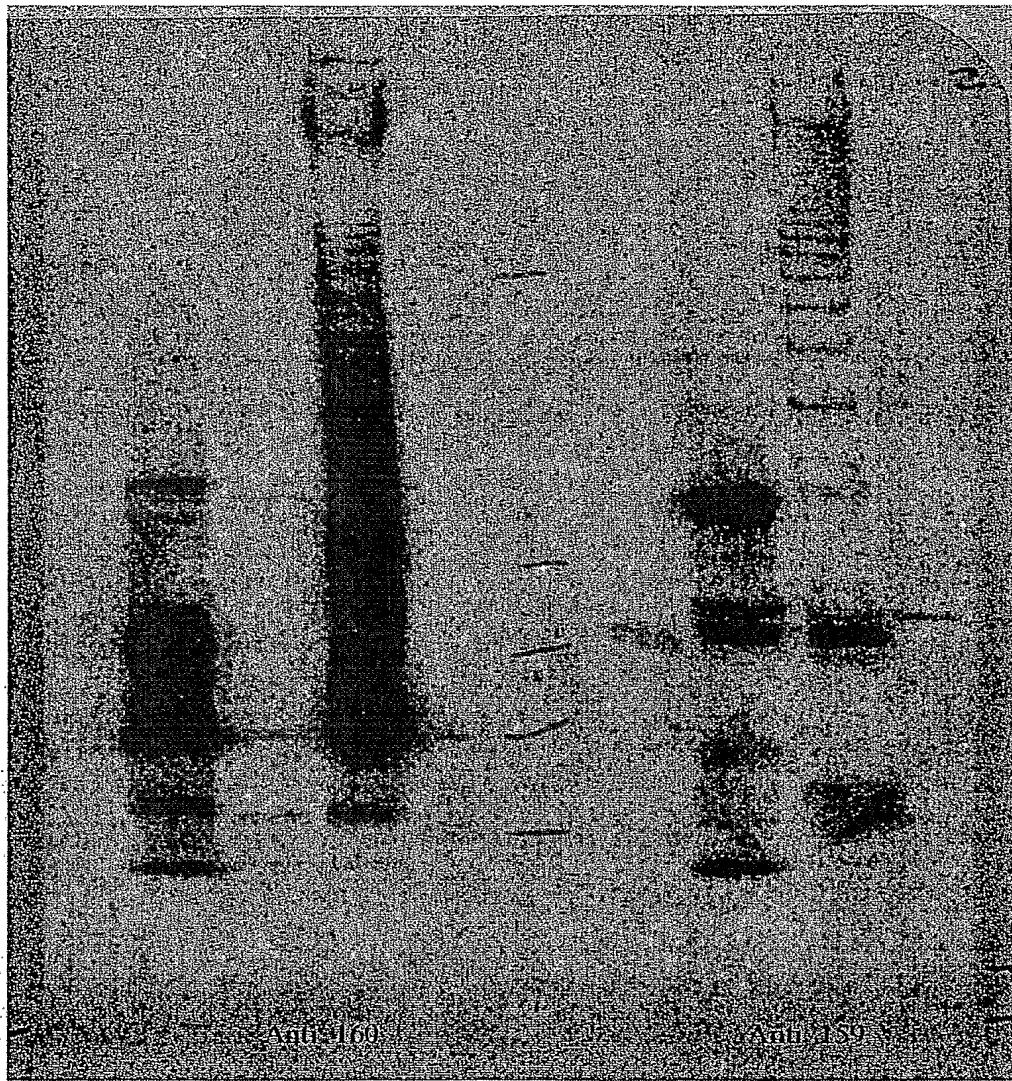
Marker

0.160

0.159

M6 3650 surf prot

M1 SF370 surf prot



M6 strain isolate 3650

Figure 111

WO 2006/078318

303/487

PCT/US2005/027239

PCT/US05/27239

M6 2724 fraz art

M6 2724 tot

0160

Marker

158

0160

0159

M6 2724 surf prot

M6 2724 tot

M1 SF370 surf prot

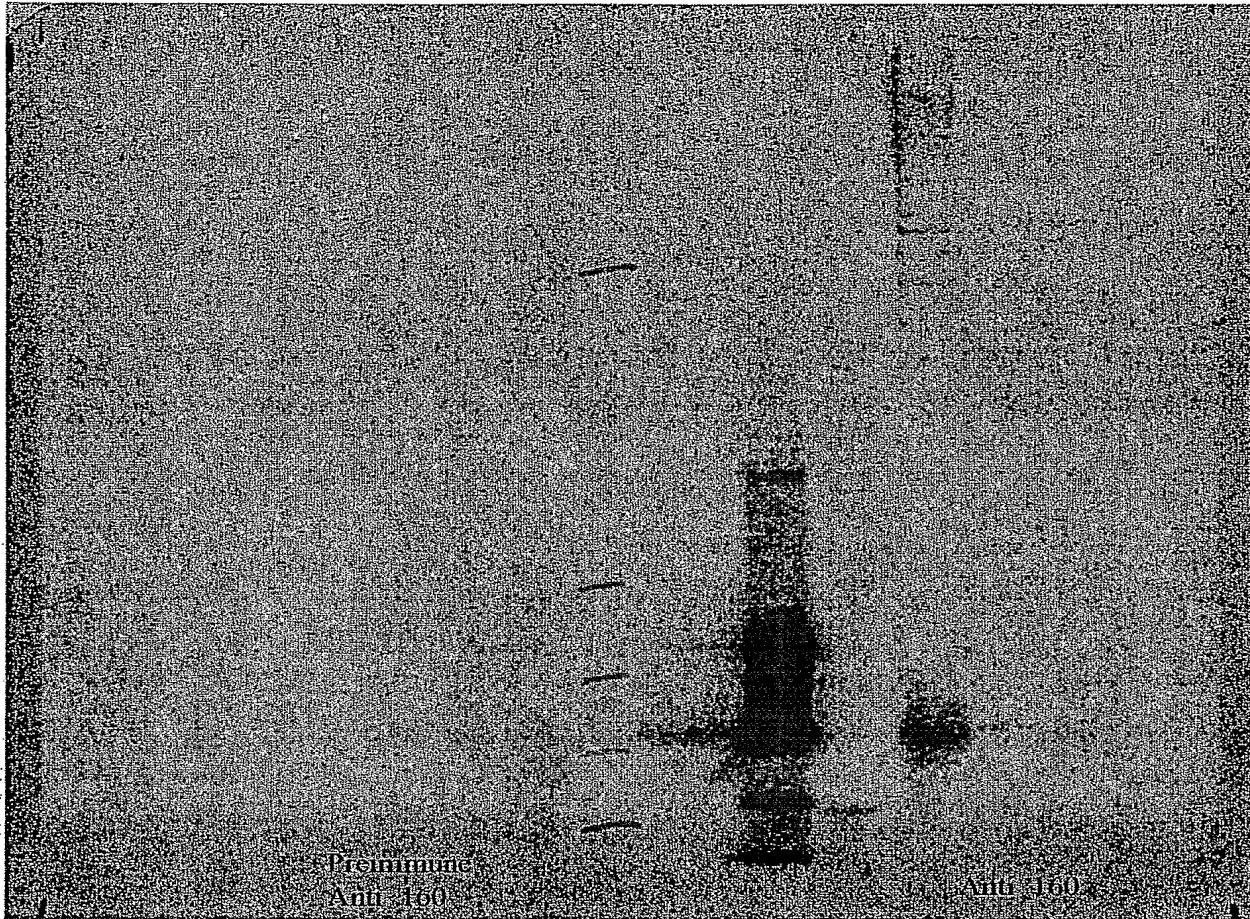


Figure 112

WO 2006/078318

PCT/US05/27239

304/487

PCT/US2005/027239

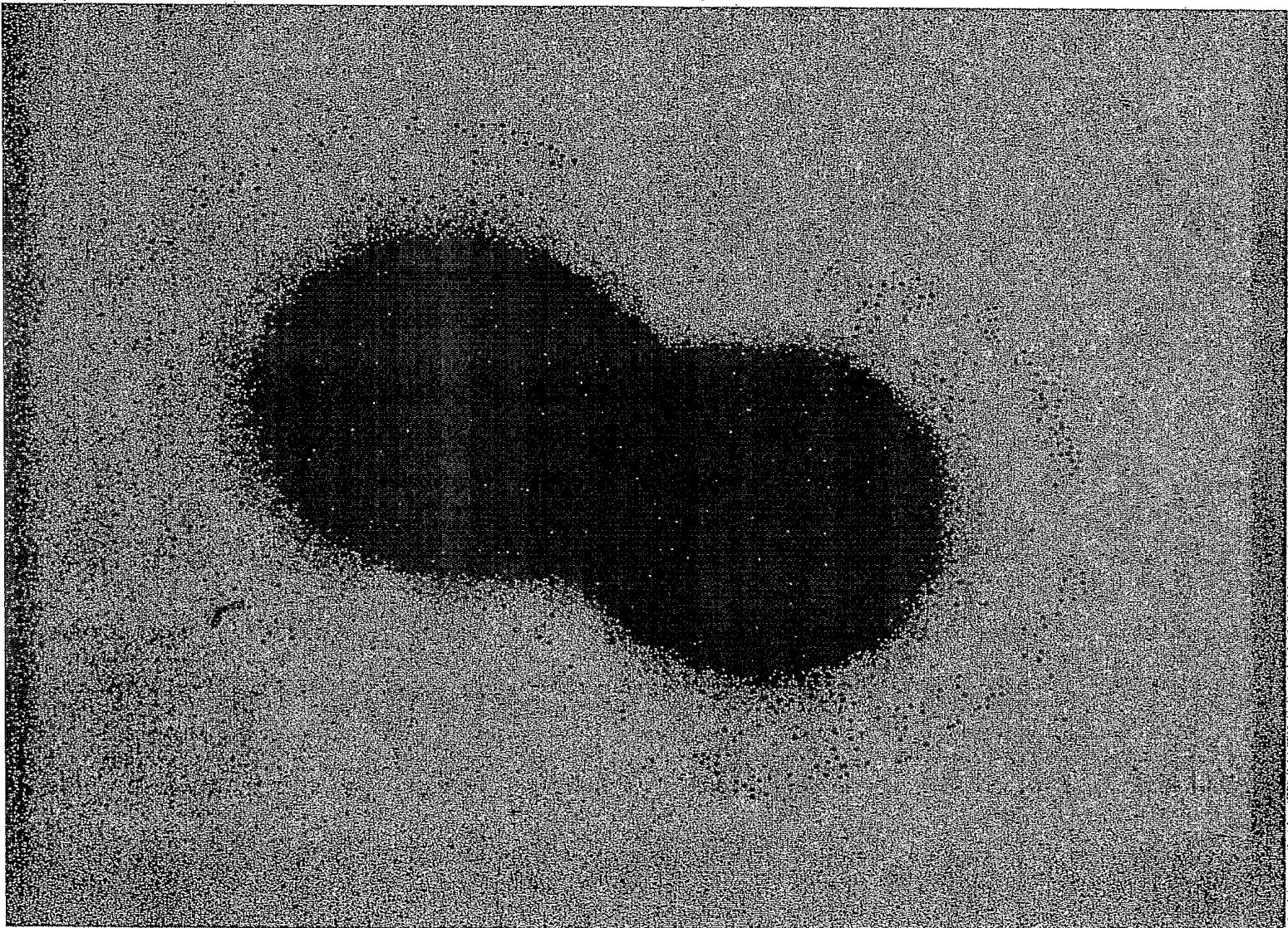


FIGURE 113

WO 2006/078318

PCT/US2005/27239

305/487

PCT/US2005/027239

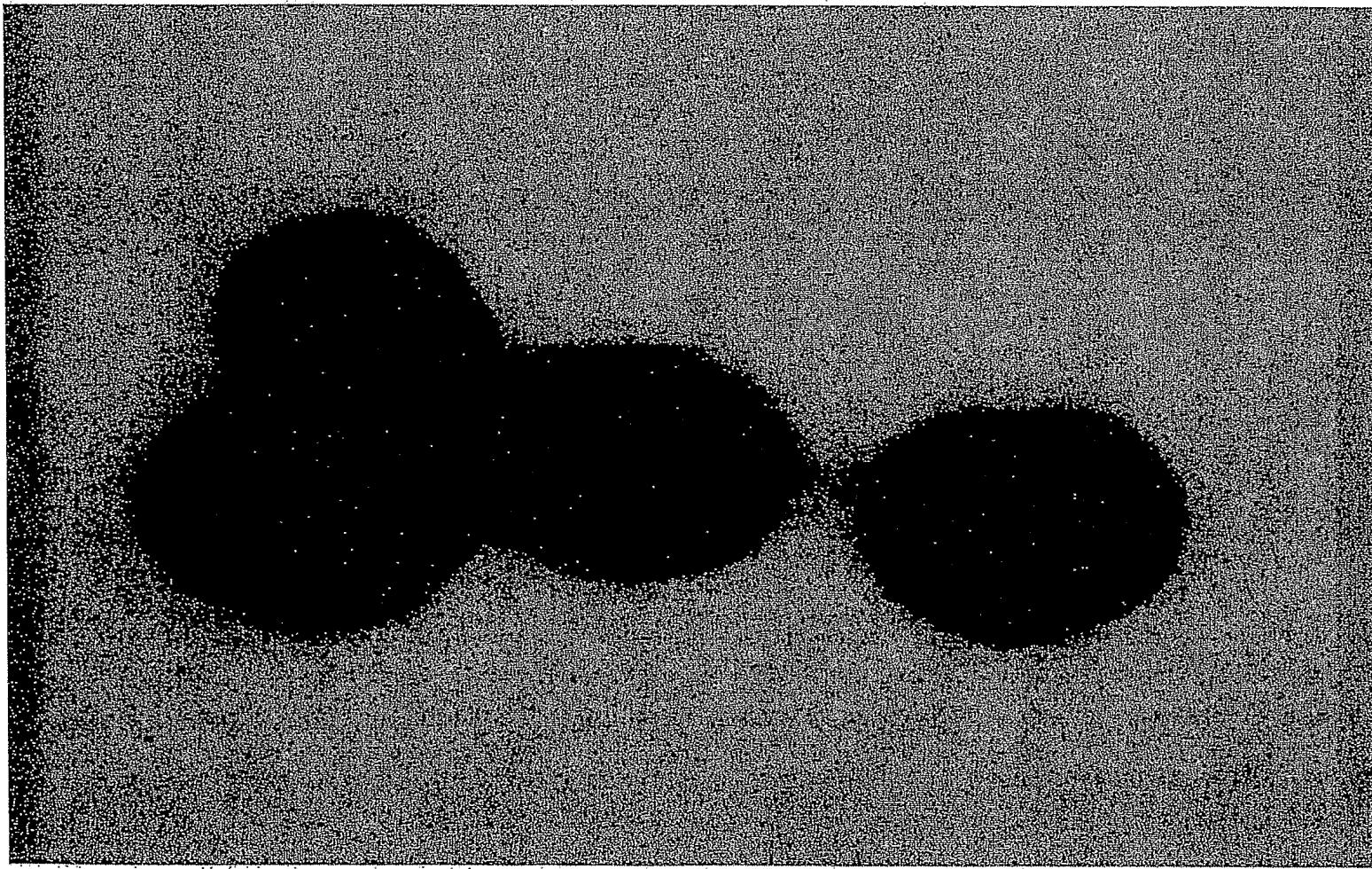


FIGURE 114

WO 2006/078318

PCT/US2005/27239 306/487

PCT/US2005/027239

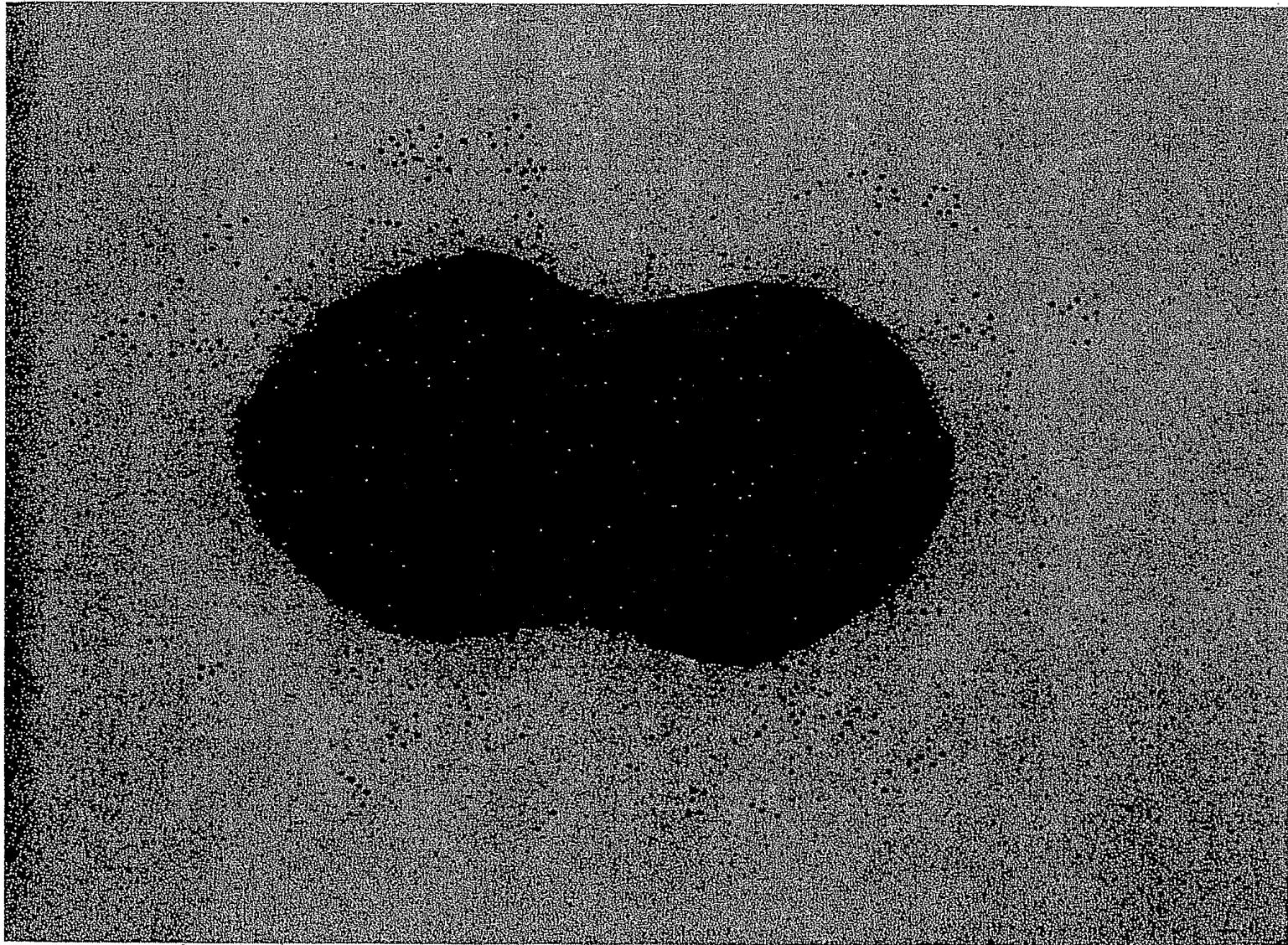


FIGURE 115

WO 2006/078318

PCT/US2005/027239

PCT/US05/27239
307/487

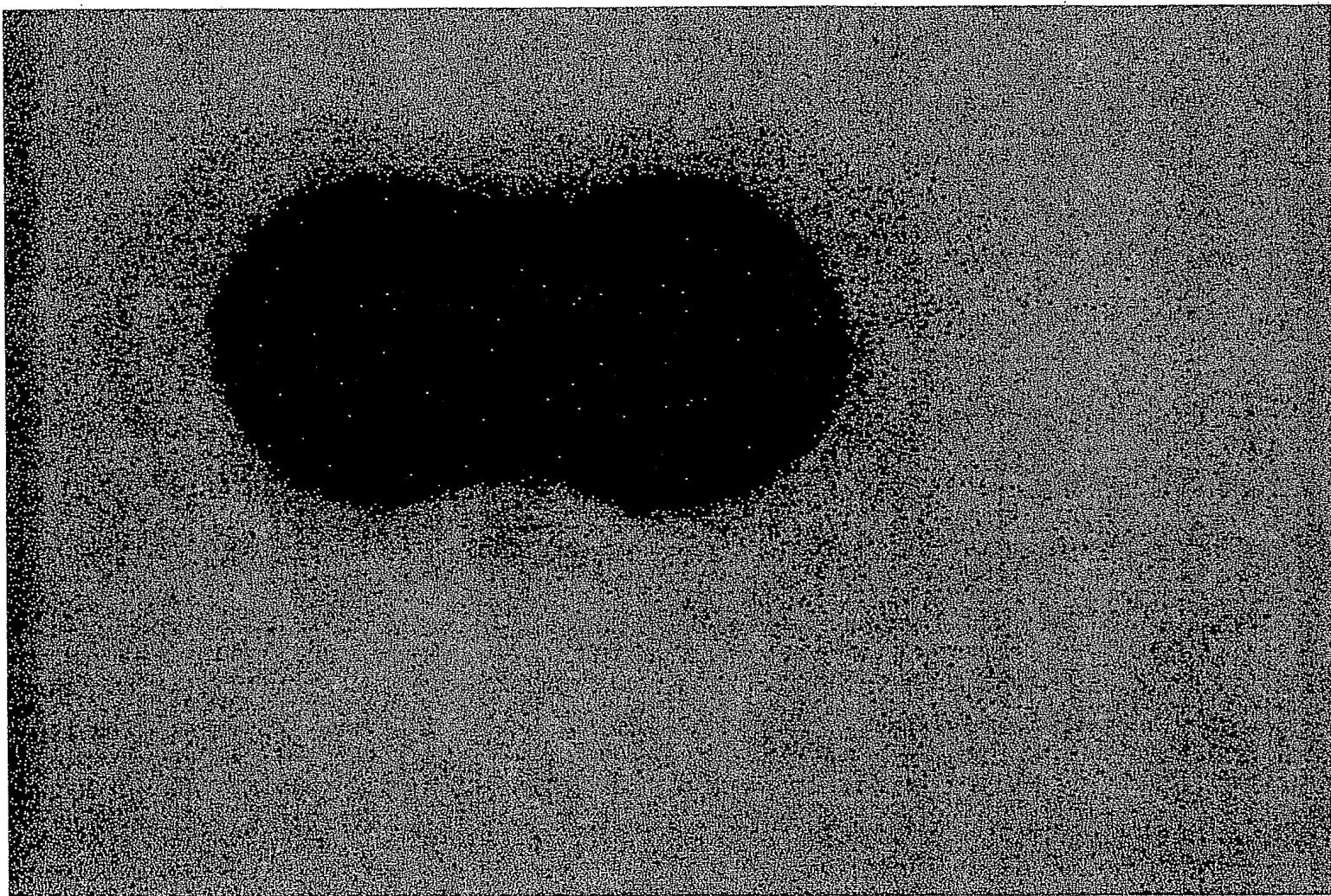


FIGURE 116

WO 2006/078318

PCT/US2005/27239

PCT/US2005/027239

308/487

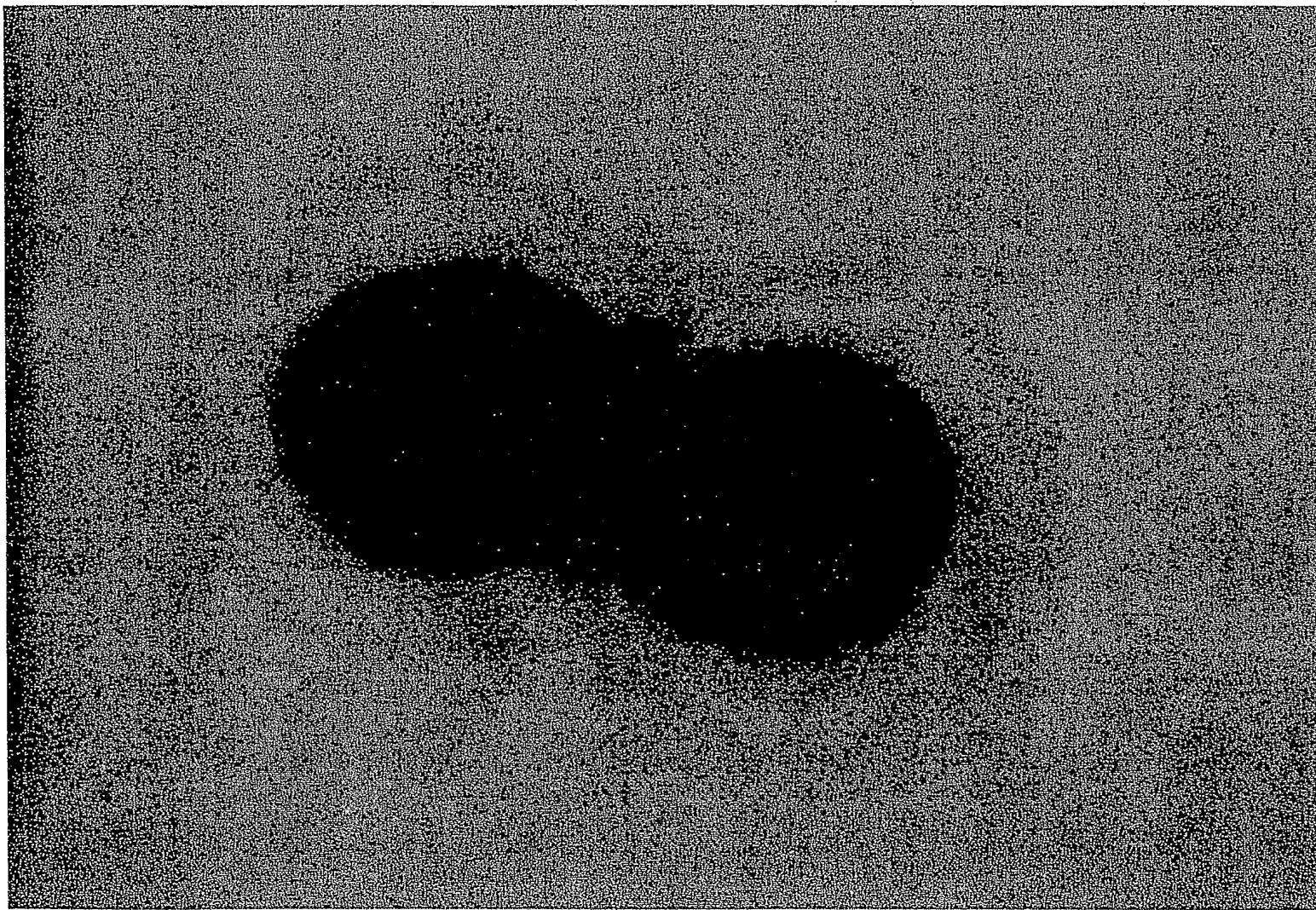


FIGURE 117

WO 2006/078318

309/487

PCT/US2005/27239

PCT/US2005/027239

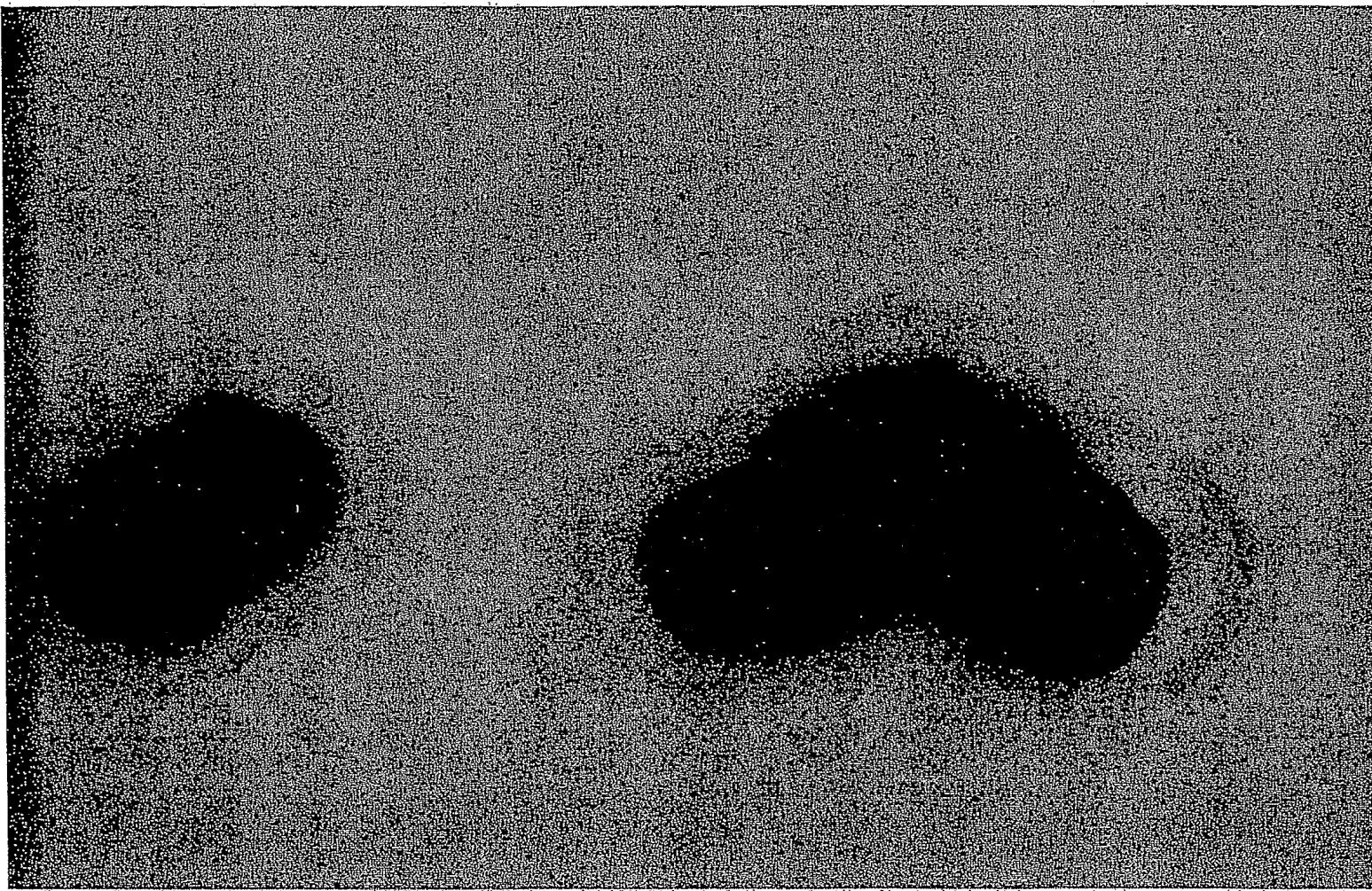


FIGURE 118

WO 2006/078318

PCT/US2005/027239

PCT/US05/27239
310/487

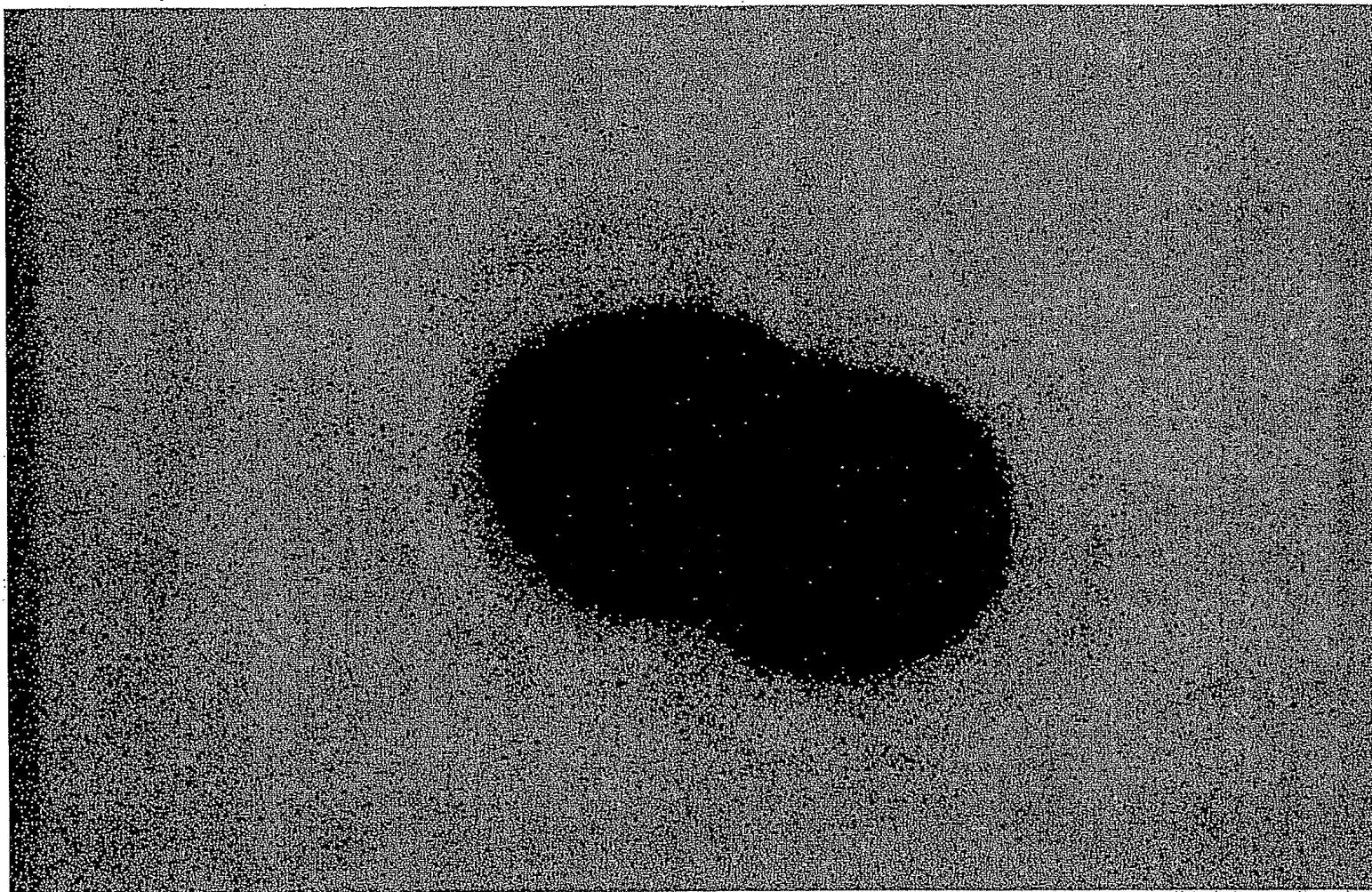


FIGURE 119

WO 2006/078318

PCT/US05/27239

PCT/US2005/027239

311/487

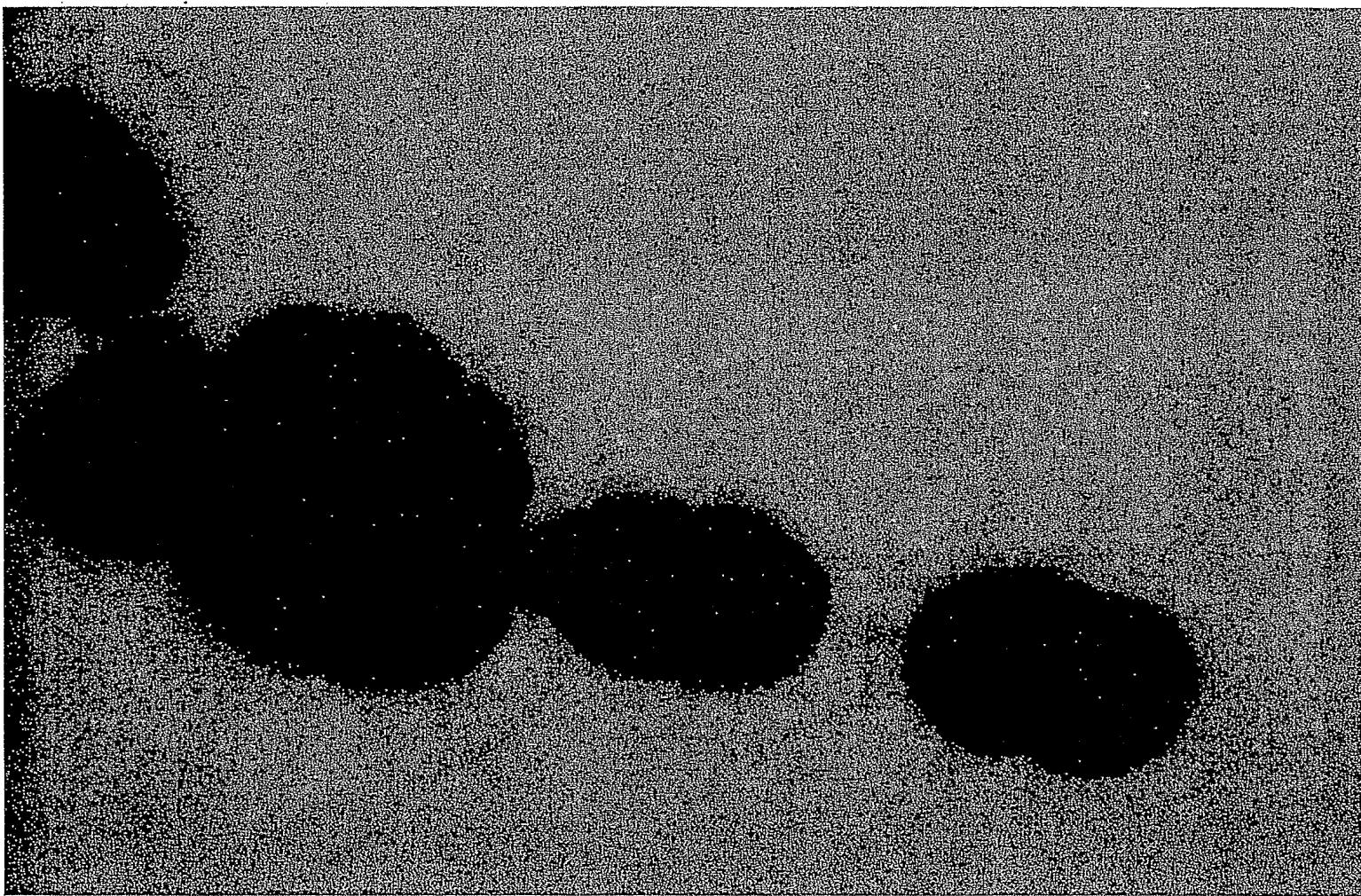


FIGURE 120

WO 2006/078318

PCT/US05/27239

PCT/US2005/027239

312/487

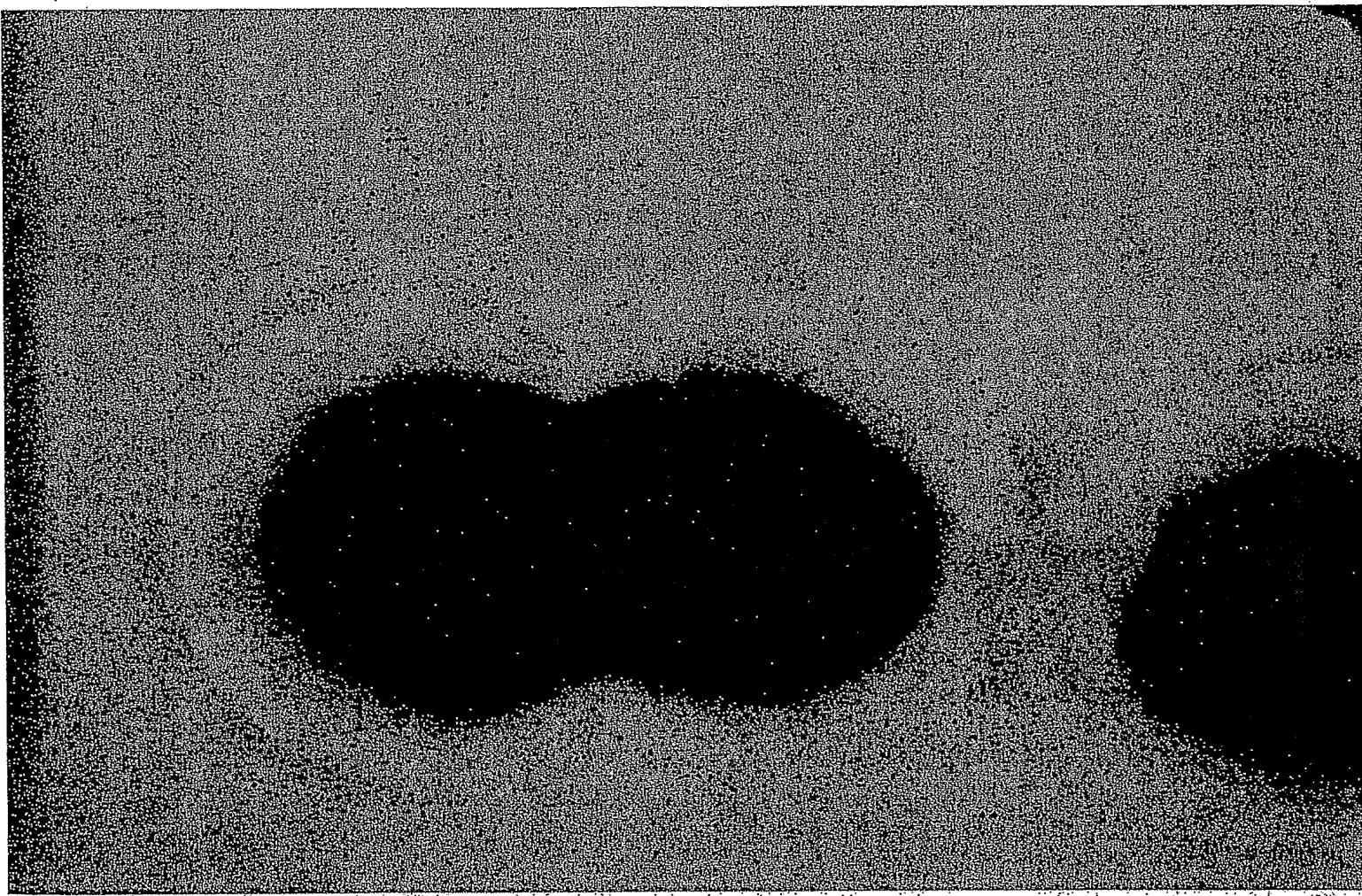


FIGURE 121

WO 2006/078318

PCT/US2005/027239

PCT/US605/27239

313/487

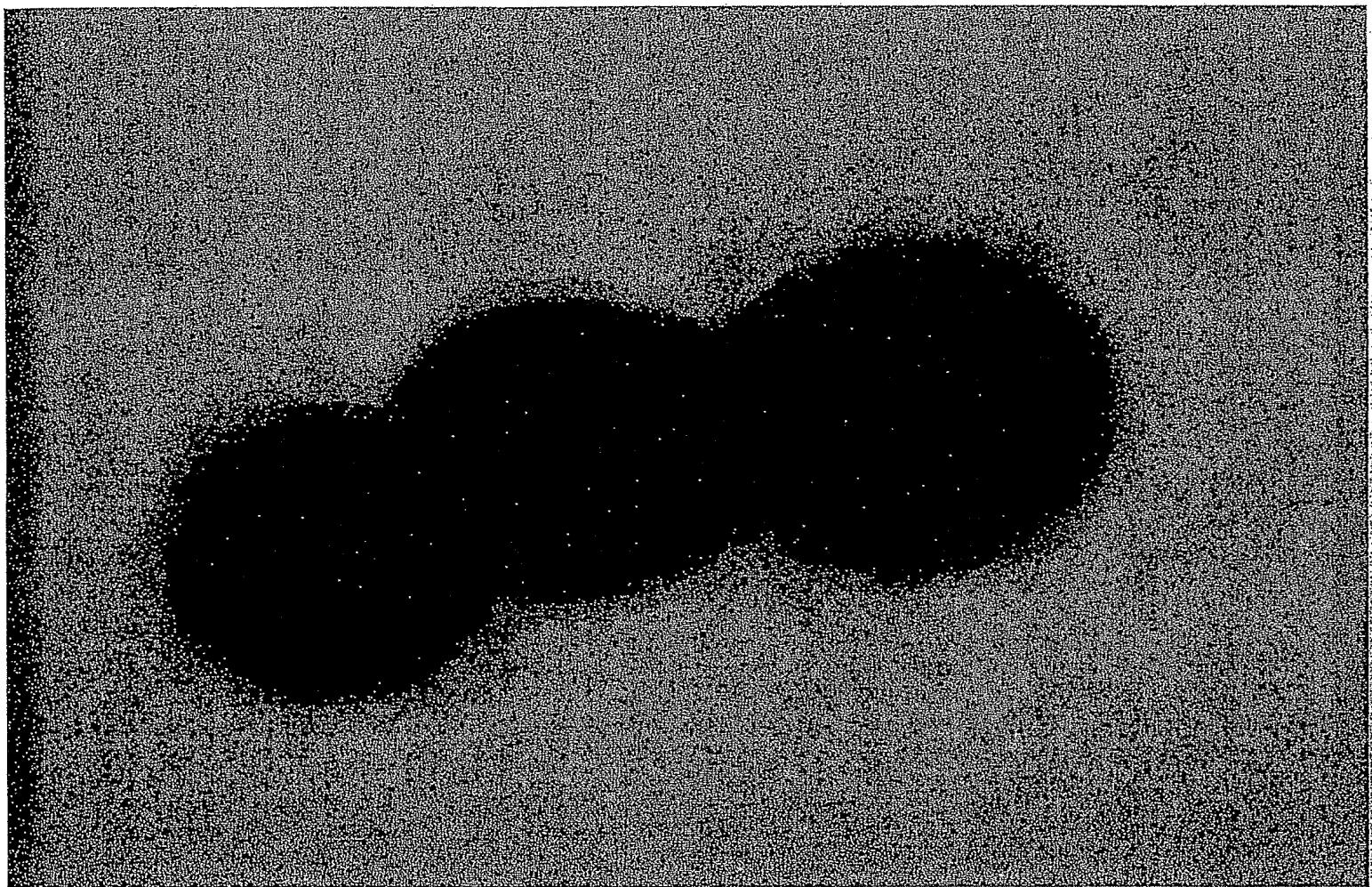


FIGURE 122

WO 2006/078318

PCT/US2005/027239

PCT/US2005/027239

314/487

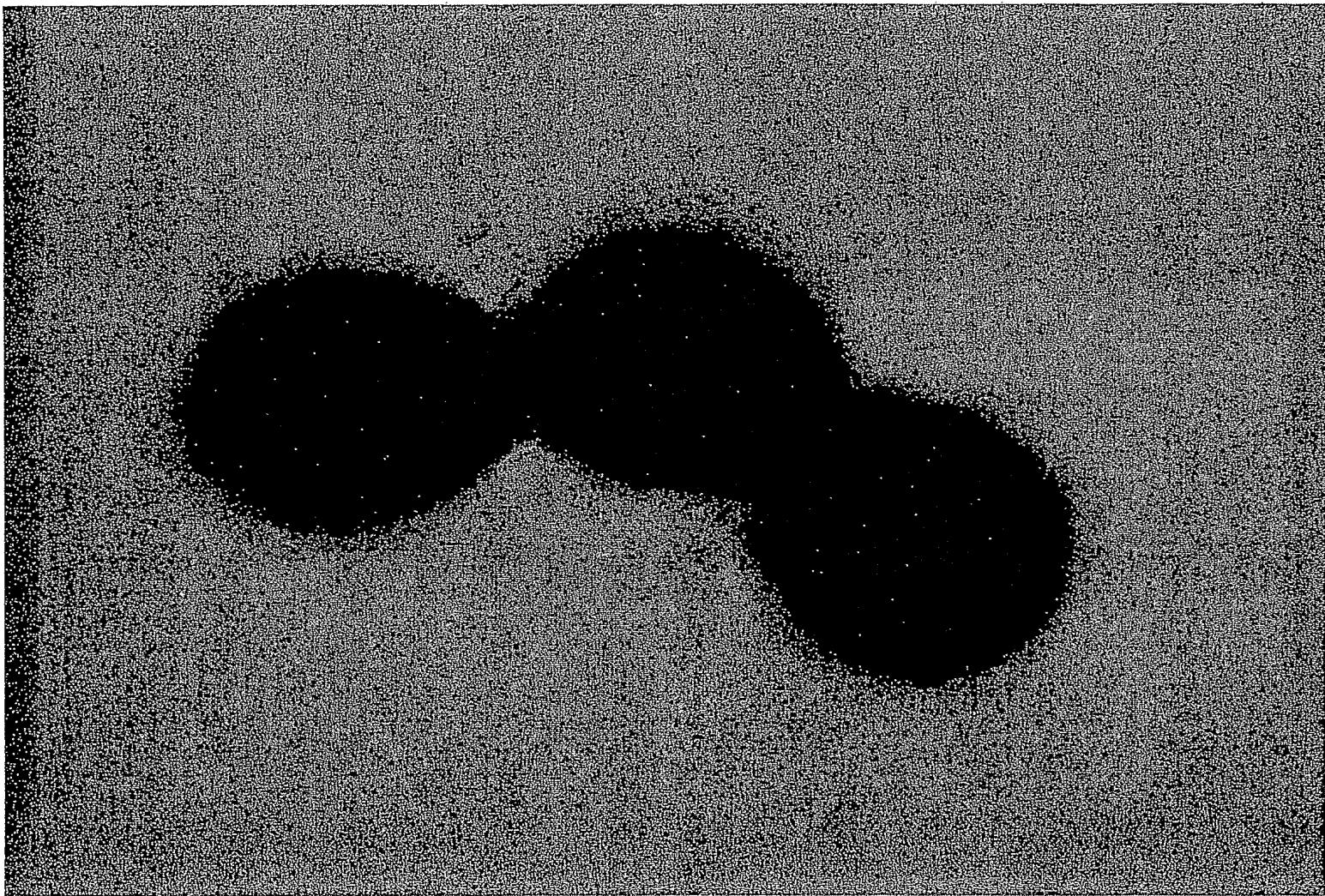


FIGURE 123

WO 2006/078318

PCT/US05/27239

PCT/US2005/027239

315/487

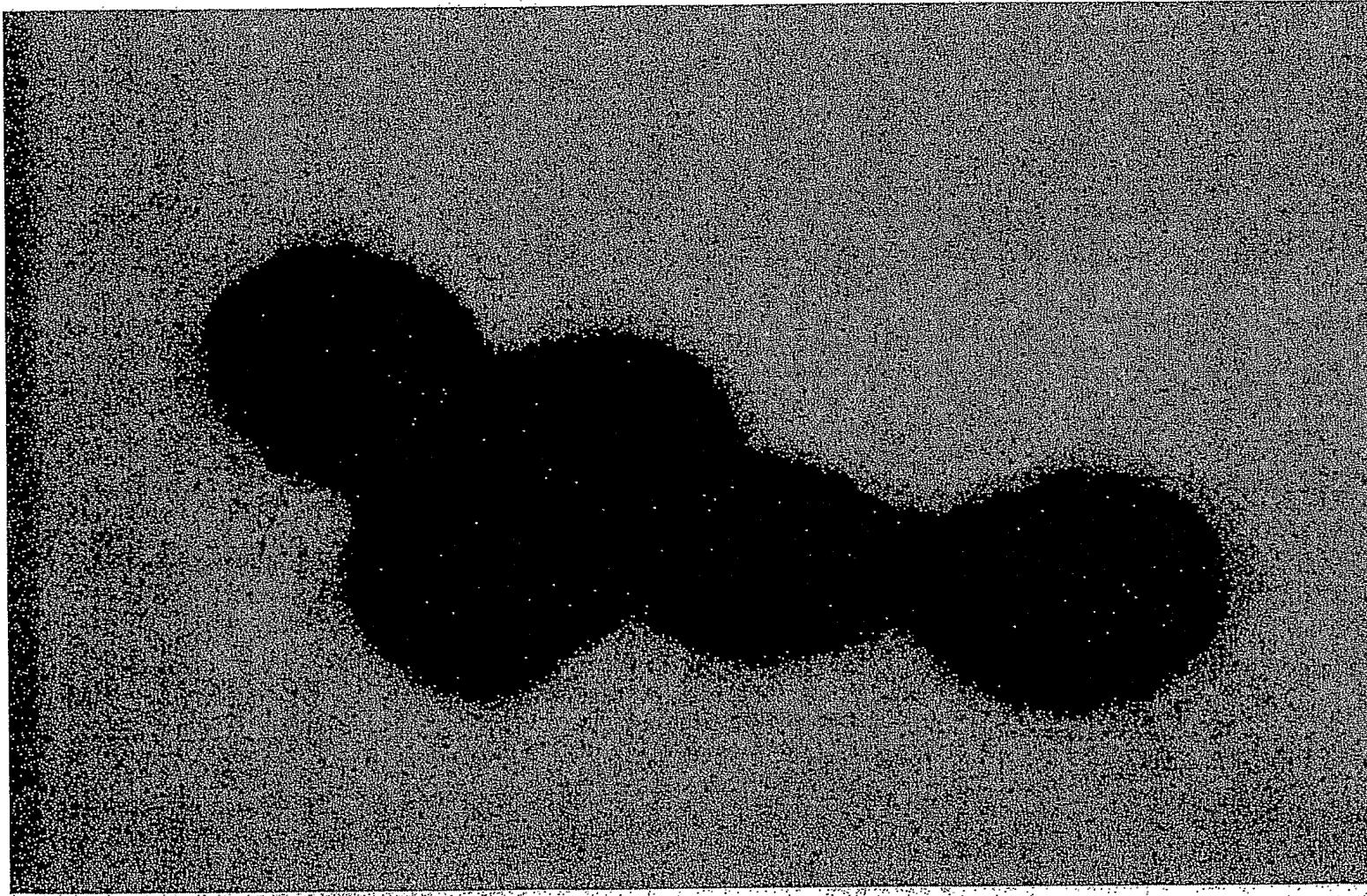


FIGURE 124

WO 2006/078318

316/487

PCT/US05/27239

PCT/US2005/027239

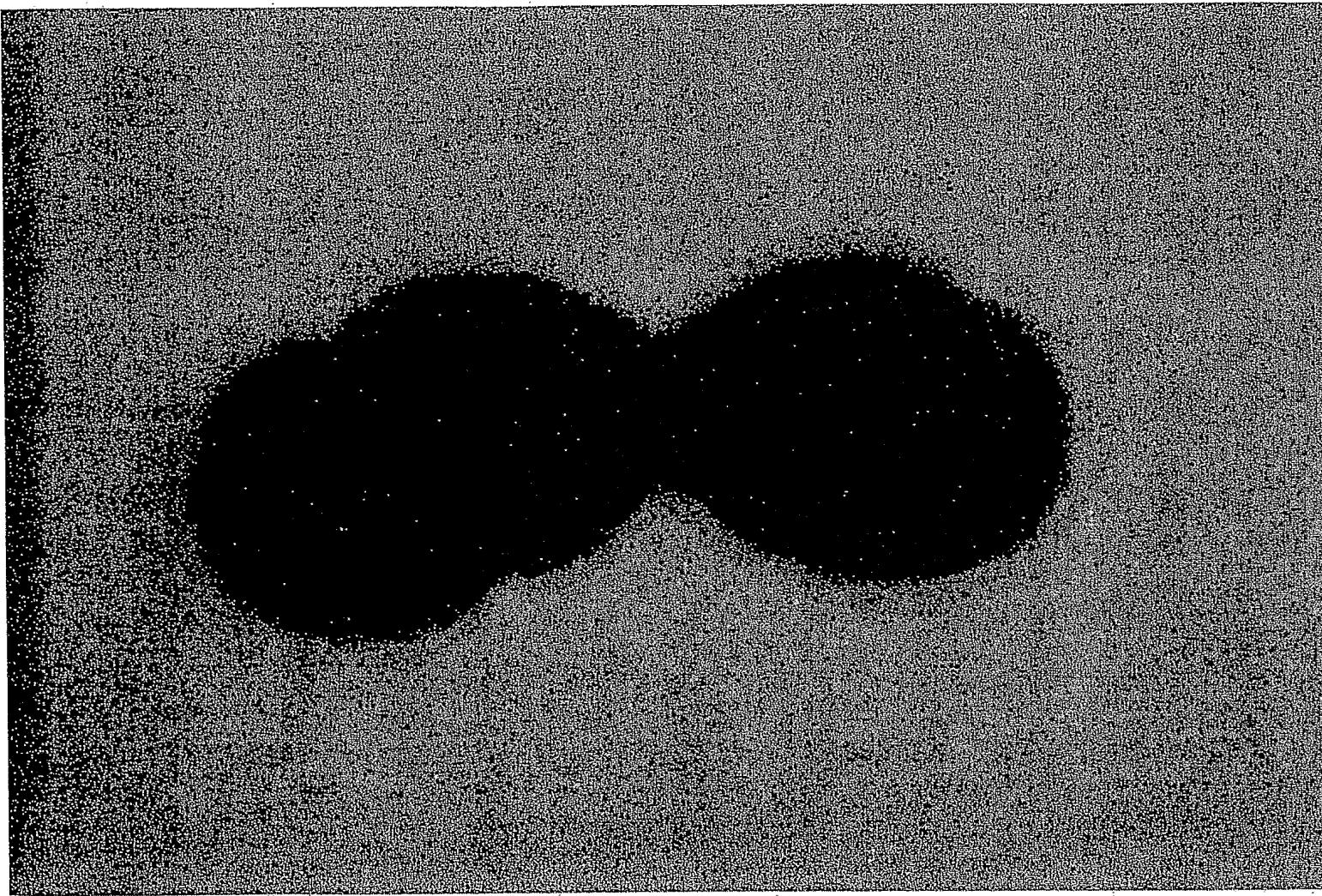


FIGURE 125

WO 2006/078318

PCT/US2005/27239

PCT/US2005/027239

317/487

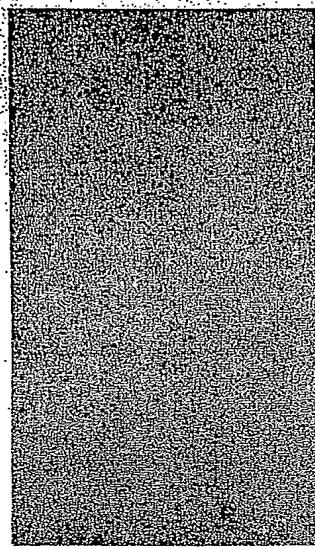


FIGURE 126

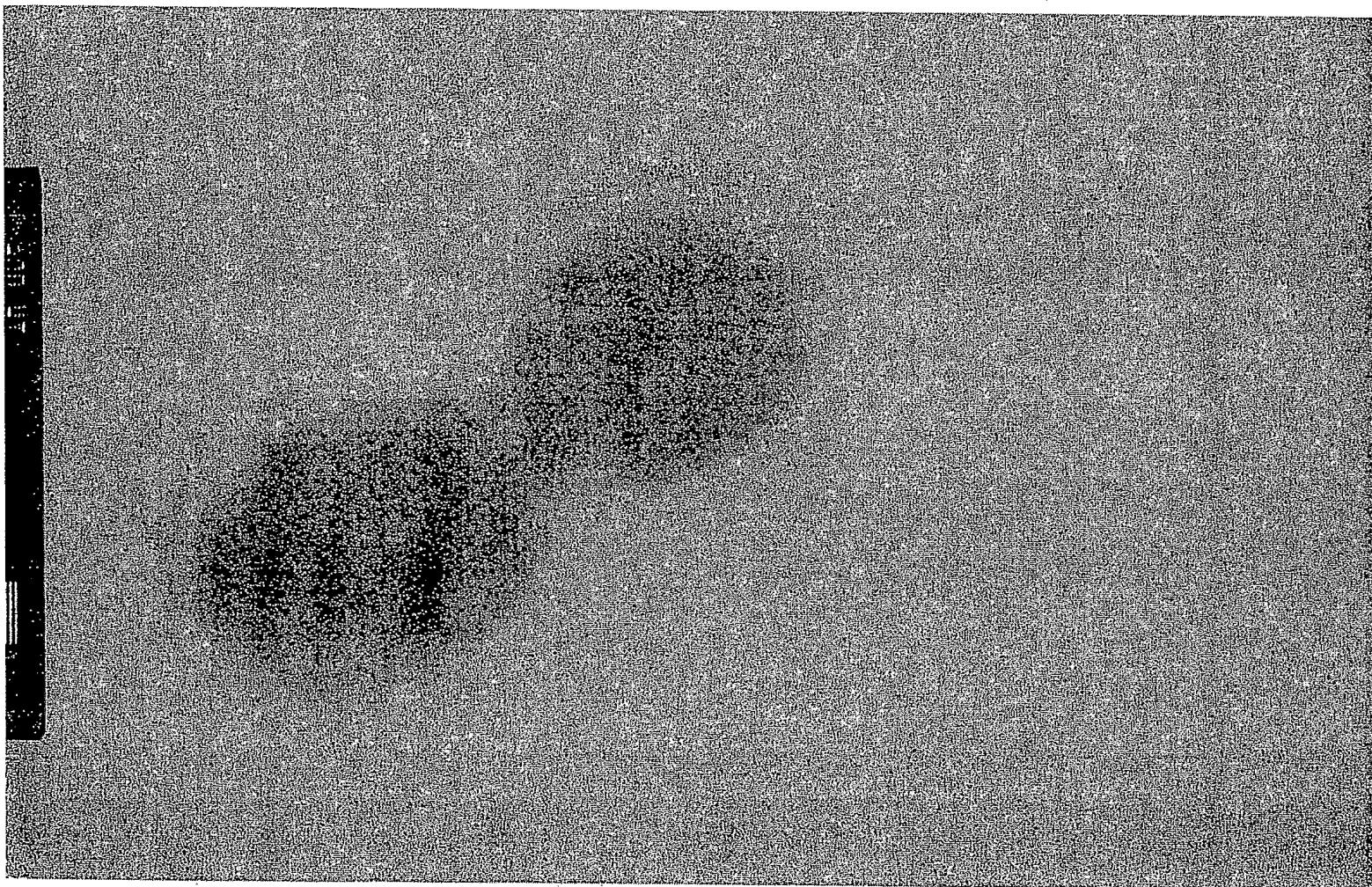
WO 2006/078318

PCT/US2005/027239

318/487

PCT/US2005/027239

Figure 127



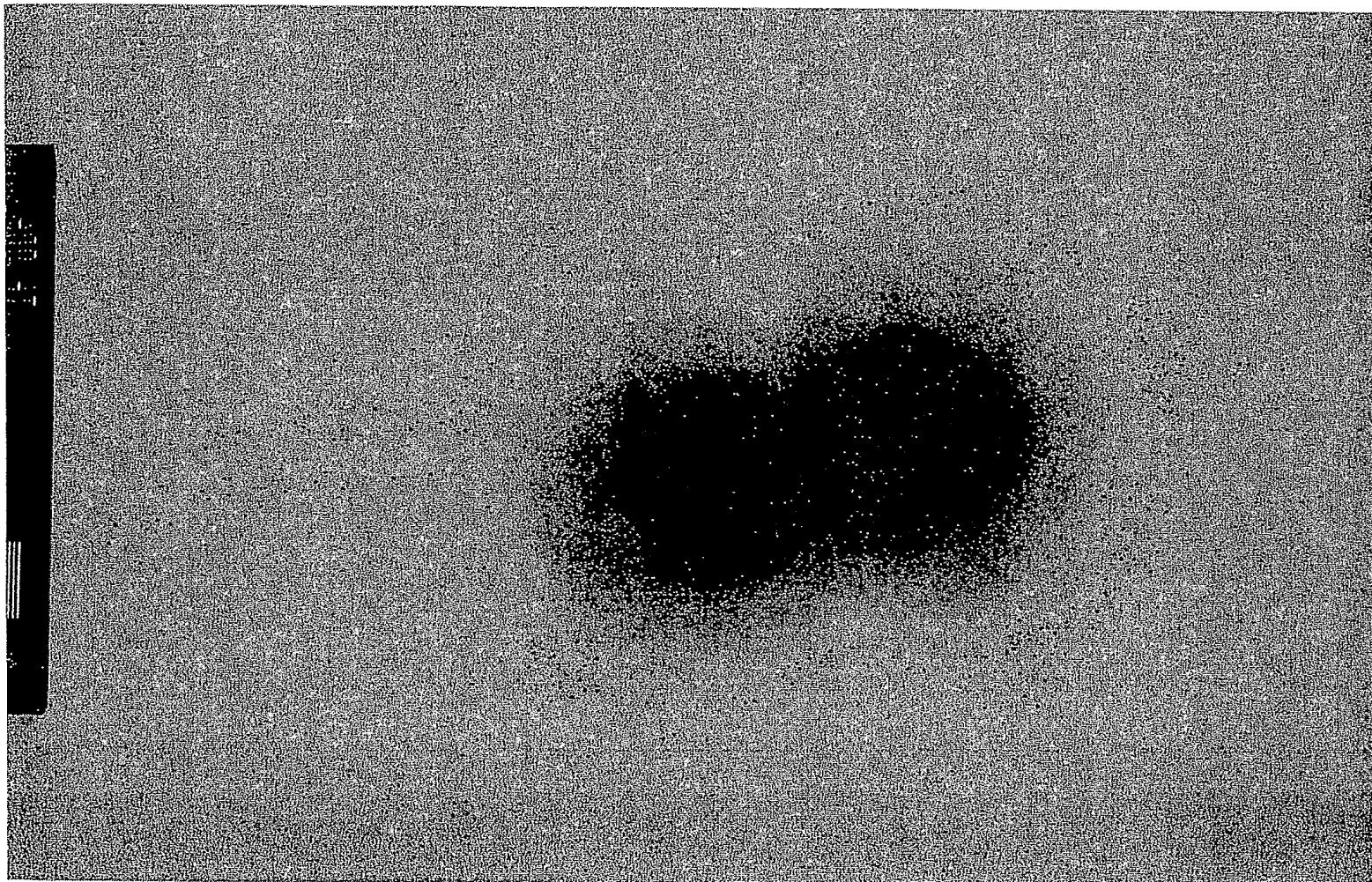
WO 2006/078318

PCT/US2005/27239

319/487

PCT/US2005/027239

Figure 128



WO 2006/078318

PCT/US2005/27239

320/487

PCT/US2005/027239

Figure 129

